Statistical communication theory and its applications

Edited by Prof. B. R. Levin

MIR PUBLISHERS MOSCOW

Contributors

Blokh E. L., D.Sc., USSR
Borodich S. V., D.Sc., USSR
Burdzeiko B. P., Eng., USSR
Chabdarov Sc. M., Cand. Sc., USSR
Finger A., Cand. Sc., GDR
Fink L. M., D.Sc., USSR
Gordos G., Prof., Hungary
Ginzburg S. A., Cand. Sc., USSR
Kalashnikov N. I., Cand. Sc., USSR
Kantor L. Ya., D.Sc., USSR
Klovsky D. D., D.Sc., USSR
Lange F. H., Corresp. Memb. of Academy, GDR
Levin B. R., D.Sc., USSR
Mirsky G. Ya., D.Sc., USSR
(continued on back flap)

Contributors

Muradyan A. G., Cand. Sc., USSR
Nemirovsky A. C., D.Sc., USSR
Szabo Cs., Cand. Sc., Hungary
Samoylenko S. I., D.Sc., USSR
Schwarz W., D.Sc., GDR
Seidler J., Corresp. Memb. of Academy, Poland
Shakhgildyan V. V., D.Sc., USSR
Shinakov Yu. S., Cand. Sc., USSR
Tsirlin V. M., Cand. Sc., USSR
Varakin L. E., D.Sc., USSR
Yakovlev V. P., D.Sc., USSR
Zigangirov K. Sh., D.Sc., USSR
Zyablov V. V., D.Sc., USSR
Zyuko A. G., D.Sc., USSR

Statistical communication theory and its applications

Статистическая теория связи и ее практические приложения

Под редакцией
Б. Р. Левина

Издательство
«Связь»
Москва

Statistical communication theory and its applications

Edited by Prof. B. R. Levin

Translated
from Russian
by
M. Edelev (Int., Chs. 1—2) and
A. Repyev (Chs. 3—9)

Mir publishers • Moscow

First published 1982
Revised from the 1979 Russian edition

The Greek Alphabet

Αα	Alpha	Ιι	Iota	Ρρ	Rho
Ββ	Beta	Κϰ	Kappa	Σσ	Sigma
Γγ	Gamma	Λλ	Lambda	Ττ	Tau
Δδ	Delta	Μμ	Mu	Υυ	Upsilon
Εε	Epsilon	Νν	Nu	Φφ	Phi
Ζζ	Zeta	Ξξ	Xi	Χχ	Chi
Ηη	Eta	Οο	Omicron	Ψψ	Psi
Θθϑ	Theta	Ππ	Pi	Ωω	Omega

На английском языке

Contents

Preface

Rather hair-splittingly, the title of this book might be "Statistical Communication Theory: Selected Topics and Applications to High-priority Problems in Telecommunications".

The idea to publish the book emerged at the First CMEA Workshop on Statistical Communication Theory and its Applications held in Pushchino near Moscow in September 1977. The participants were scientists and engineers from Bulgaria, GDR, Hungary, Poland and the Soviet Union. Three Corresponding Members of the Academies, 39 Professors and Doctors of Science, 60 Candidates of Science represented the leading science centers and universities of the member-countries. The main objectives of the Workshop were to listen to reviewing reports on the major research areas of the theory, to exchange applicational experience, and to work out recommendations concerning the strategy of further research and applications.

The reviewing reports delivered at the Workshop make up the body of the book. Yet, it should not be viewed as a collection of abridged papers (nonabridged form would require a volume twice as large). The authors agreed that the contributions to the book should be selected, arranged and presented according to a preset structure and methodology, that unified terms and symbols should be used throughout the text as far as possible, and, finally, that the book should have a common list of references. The editor working in contact with the authors tried to make his best in achieving these not unlaborious goals.

The book has an introduction, nine chapters, and a list of references with more than 500 entries published mainly in the last decade.

The introduction outlines the methodologic principles and gives a historical survey of statistical communication theory paralleling the sequence of the papers in the book. Basically, every chapter discusses theoretical issues within the framework of applications. Yet, to be more specific, the first five chapters focus on theoretical aspects, and the last four on applications.

Chapter 1 is devoted to stochastic models of signals, interference and communication channels, and to the measurement of statistical parameters of these objects.

Chapter 2 presents an original approach to the description of stochastic system behavior, evaluates interesting parallels between the description of the deterministic and stochastic systems. For the latter, two ways of presentation are given.

Chapter 3 treats methods of coding and decoding, with main emphasis on difficulties in their realization.

Chapter 4 deals with the theory and application of large time-bandwidth product signals.

Chapter 5 illustrates what can be done for the synthesis of information systems under uncertainty by statistical methods.

Chapter 6 gives an analysis of interference and crosstalk in multichannel radio communication systems.

Chapter 7 shows how statistical communication theory is embodied in practical developments exemplified by satellite, troposcatter and optical communication systems and electromagnetic compatibility of radio systems.

The last two chapters discuss computer-communication networks.

On the whole, the book covers a significant part of modern statistical communication theory along with vast areas of its application in telecommunications, giving the state of the art by the end of 1977.

B. R. Levin

The authors are deeply grateful to Prof. I. E. Efimov, Chairman of the Organizing Committee for the Pushchino Workshop, Rector of the Moscow Institute of Telecommunication, and Prof. V. I. Siforov, director of the Institute of Information Transmission for their encouragement to the idea to publish this work and for their support to its realization.

Contributors

F. H. Lange, Corresponding Member of the GDR Academy of Sciences (Sec. I.3)
B. R. Levin, D. Sc. (Tech.), Prof. (Introduction—but for Sec. I.3; Secs. 5.1-5.6)
Yu. S. Shinakov, Cand. Sc. (Tech.), Assist. Prof. (Secs. 1.1-1.3, 1.6)
D. D. Klovsky, D. Sc. (Tech.), Prof. (Sec. 1.4)
G. Ya. Mirsky, D. Sc. (Tech.), Prof. (Sec. 1.5)
W. Schwarz, D. Sc. (Tech.), Prof., Higher Technical School, Mittweida GDR (Chap. 2)
E. L. Blokh, D. Sc. (Phys.), Prof. (Sec. 3.1)
V. V. Zyablov, D. Sc. (Tech.) (Sec. 3.1)
K. Sh. Zigangirov, D. Sc. (Tech.) (Sec. 3.2)
L. M. Fink, D. Sc. (Tech.), Prof. (Sec. 3.3)
A. G. Zyuko, D. Sc. (Tech.), Prof. (Sec. 3.4)
L. E. Varakin, D. Sc. (Tech.), Prof. (Secs. 4.1-4.3, 4.5)
A. Finger, Cand. Sc. (Tech.), Technical University, Dresden, GDR (Sec. 4.4)
V. V. Shakhgildyan, D. Sc. (Tech.), Prof. (Sec. 5.7)
B. P. Burdzeiko, Eng. (Sec. 5.7)
Sh. M. Chabdarov, Cand. Sc. (Tech.), Assist. Prof. (Sec. 5.8)
S. V. Borodich, D. Sc. (Tech.), Prof. (Secs. 6.1-6.2)
G. Gordos, Prof., Technical University of Budapest, Hungary (Sec. 6.3)
L. Ya. Kantor, D. Sc. (Tech.) (Sec. 7.1)
V. M. Tsirlin, Cand. Sc. (Tech.) (Sec. 7.1)
V. P. Yakovlev, D. Sc. (Tech.) (Sec. 7.2)
A. S. Nemirovsky, D. Sc. (Tech.) Prof. (Sec. 7.3)
N. I. Kalashnikov, Cand. Sc. (Tech.), Prof. (Sec. 7.4)
S. A. Ginzburg, Cand. Sc. (Tech.) (Sec. 7.5)
A. G. Muradyan, Cand. Sc. (Tech.), Assist. Prof. (Sec. 7.5)
J. Seidler, Corresponding Member of the Polish Academy of Sciences (Secs. 8.1-8.3)
Cs. Szabo, Cand. Sc. (Tech.), Technical University of Budapest, Hungary (Sec. 8.4)
S. I. Samoylenko, D. Sc. (Tech.) (Chap. 9)

List of Principal Symbols

A = event
a = amplitude of a signal
$A(t)$ = quadrature component of a narrowband process
B = time-bandwidth product of a signal
$B(t)$ = quadrature component of a narrowband process
C^1 = space of continuous functions
D = signal distance
E = experiment; energy in a signal
$\mathrm{E}[\ ]$ = expected value of the quantity in brackets
F = spectrum bandwidth
$F(x)$ = normal (Gaussian) distribution function
$F(\omega)$ = power spectrum
$F_1(x, t)$ = one-dimensional probability distribution function
$F(\mathbf{x}_1^n, \mathbf{t}_1^n)$ = joint probability distribution function
f_0 = carrier frequency
GF = Galois field
h = signal-to-noise ratio (S/N)
$h(u, v)$ = impulse response of a linear system
$I_0(x)$ = zero-order modified Bessel function
$\mathbf{I}$ = identity (unit) matrix
$K(u, v)$ = covariance function
$\mathbf{K}$ = covariance matrix
$K(i\omega)$, $K(s)$ = transfer function of a linear network
m = mean value of a random variable
n = size of a sample; code word length
$n(t)$ = white noise
N = number of trials; number of scatter elements in multipath propagation
N_0 = spectral density of white noise
p = probability
$\mathrm{P}\{A\}$ = probability of an event A
P_s = power of a signal
P_n = power of noise
R = transmission rate
$R(t)$ = normalized autocorrelation function
$s(t)$ = desired signal
T = duration of an observation inverval; signal length; set of time points
t, τ = time variables
$w(\mathbf{x}_1^n, \mathbf{t}_1^n)$ = joint probability density
$W(t)$ = Wiener process
$\mathscr{X}$ = ensemble of realizations

x = threshold value; an independent variable
$x(t)$ = realization of a random process
$\mathbf{X}$ = set of input values
$X(t)$ = process
$Y(t)$ = process
$\mathcal{Y}$ = set of conditions
$\mathbf{Y}$ = set of output values
z = state-variable
$\delta(\tau)$ = Dirac delta function
Δ = frequency band
$\hat{\theta}$ = estimate of a parameter θ
π = conditional probability matrix
σ^2 = variance
Φ = phase of a signal
φ = phase
τ = delay time
Ω, ω = angular frequency
ξ, η, ζ, ν = random variables
$\xi(t)$ = interference, noise
$\in$ = "belongs in" or "falls in"
$\triangleq$ = equal by definition
$\forall$ = "for all"
$\operatorname{sgn}(\cdot)$ = sign function, equal to $+1$ for positive argument and to -1 for negative argument

Abbreviations

Modulation and Multiplexing

ADM = adaptive delta modulation
AM = amplitude modulation
DPSK = differential phase-shift keying
FDM = frequency-division multiplex
FM = frequency modulation
FSK = frequency-shift keying
PAM = pulse-amplitude modulation
PCM = pulse-code modulation
PM = phase modulation
PPM = pulse-position modulation
PSK = phase-shift keying
PWM = pulse-width modulation
SSB = single-sideband (AM)
TDM = time division multiplex

Organisations

CCIR = Comité Consultatif International de Radio
CCITT = Comité Consultatif International de Télégraphie et Téléphonie
CMEA = Council for Mutual Economic Assistance

Others

ACF = autocorrelation function
CCF = cross-correlation function
CVP = channel with variable parameters
DFT = discrete Fourier transform
DSC = discrete symmetric memoryless channel
FFT = fast Fourier transform
IF = intermediate frequency
LSC = linear stochastic channel
PLL = phase-lock loop
RF = radio frequency
RSC = random structure channel

Introduction

A Historical Survey of Statistical Communication Theory

I.1. Pioneering the Statistical Communication Theory

In 1947, V. A. Kotel'nikov, Vice-President of the USSR Academy of Sciences at this writing, presented to the Moscow Power Institute the dissertation headed "A theory of potential noise immunity". In the submitted thesis he formulated the objectives of optimal statistical synthesis of communication receivers in the presently known form. From the new positions he analyzed various communication systems to found out the bounds for feasible methods of modulation [151].

Over a year later, the Bell System Technical Journal published Claude Shannon's papers "A mathematical theory of communication" [478] where he set forth his two famous theorems. The first is on message source encoding to offset redundancy and the second on the maximum amount of information which by a suitable coding procedure can be transmitted through the discrete noisy channel with arbitrary small error rate. This work initiated coding theory as another active area in communications research.

The fundamental results by Kotel'nikov and Shannon have been widely recognized as basic contribution into the building of statistical communication theory. Three past decades have seen substantial progress in the methods of the theory and their applications in the research and development work.

I.2. General Methodology

Consider shortly the historical backgrounds to statistical communication theory which may serve another illustration of the knowledge continuity and interrelation principles in science.

Philosophical concepts. Since the time of Newton and Laplace physical processes were studied predominantly using deterministic principles. Although skilful and consistent in their use, they, how-

ever, could not embrace all the aspects of real phenomena. So, for example, the methods of classical mechanics, which had seemed adequate to sufficiently describe microcosmic processes, had to give up their place to a probabilistic approach that in the cource of an elaborated study proved a powerful alternative to the deterministic techniques.

At the turn of this century the well-balanced structure of determinism in physics suffered a first deep break. In 1902, American theoretical physisist J. Willard Gibbs published his work on the general principles of statistical mechanics [78]. These principles paved the way for the quantum theory of elementary particles, which is based on the known principle to determine only the probability of an experimental outcome. Max Born, one of the founders of quantum mechanics, wrote that casuality of classical physics had always been interpreted as determinism (even by Immanuael Kant) and that new quantum mechanics turned out to be statistical rather than deterministic, and its success in all fields of physics was a fact. [31].

First applied at a microscopic level, the probabilistic and statistical concepts of physics penetrated deep into the natural sciences to interprete manifold macroscopic phenomena and to describe them clearly and exactly even in continuous media. In the 1940s, these concepts gained a reliable foothold in radio engineering and communication theory. It became obvious that in designing a communication system one had to reckon with the effect of noise although it could not be exactly predicted and deterministically corrected due to its stochastic nature. On the other hand, transmission of information in a communication system is sensible only when the transmitted message is not known a priory to the recepient. This pragmatic approach led to the choice of a stochastic model for the message source and communication channel; it also suggested that instead of an individual signal an ensemble of signals should be considered using a probability measure specified on this ensemble.

Following statistical mechanics and statistical physics, the middle of the century saw the advent of statistical radio physics, statistical radio engineering, statistical communication theory, and theory of information. The concept of indeterminism is that joins all of them.

Mathematics involved. To deduce the quantitative characteristics of the studied process the scientist needs an adequate body of mathematics. Now that the deterministic principles had to be abandoned a need arose to expand the mathematical methods to include first of all those of theory of probability and mathematical statistics. Gaining new applications, however, mathematics itself had to adapt to the new needs. For example, random process theory branched off probability theory, and decision theory branched off mathematical statistics.

These disciplines were given a far-reaching impetus by several outstanding mathematicians. In the early 1930s, Khinchin and Wiener developed the theory of harmonic analysis of random functions. So, the Wiener-Khinchin theorem relates the autocorrelation function and power spectrum of a wide-sense stationary random process. In 1939, Kolmogorov published—first abroad [417], then in the USSR [140]— a paper which laid a foundation for further work on random process filtering.

In 1942, Wiener wrote a report on random process filtering, which was made public in 1949 [500]. At this time he was already aware of the Kolmogorov's findings. Later in his book 'I am a Mathematician' Wiener wrote, "My research at this time received a ready reception in Russia and was in close relation with the work of some of the Russian mathematicians, although I had never met any of them nor, I believe, ever been in correspondence with them. Khinchin and Kolmogorov, the two chief Russian exponents of the theory of probability, have long been involved in the same field in which I was working. For more than twenty years, we have been on one another's heels; either they had proved a theorem which I was about to prove, or I had been ahead of them by the narrowest of margine. This contact between our work came not from any definite program on my part, nor I believe, from any of theirs but was due to the fact that we had come into our greatest activity at about the same time, with about the same intellectual equipment."

Problems and objectives. Any classification arranges things according to some attributes. With regard to problems and objectives of statistical communication theory this may be done with recourse to a block diagram where a communication system is presented as a sequence of "black boxes" running from source to sink (see, e.g. [96]).

Two basic groups of objectives are those of analysis and synthesis. In analysis, the objective is to study and deduce the system or subsystem of interest assuming that the system algorithm is given along with the probability distribution of signals and noise on the input side. In synthesis, the objective is to develop a system or subsystem algorithm complying with a specified performance criterion.

Synthesis may be based on a complete a priori information, when the probability distributions of the signals and noise and the additional constraints are known. Alternatively, it may be attempted under conditions of uncertainty when some of the a priori data used as the performance criterion are not known. For a more detailed classification, one or more "black boxes" must be isolated from the block diagram. Then several groups of problems and objectives may be visualized in statistical communication theory:

model buildings for signals, noise, and communication channels:
coding and decoding;
multiplexing in a multichannel system;

choice of signal waveforms and modulation methods;
analysis of noise, interference and distortion in a system;
choice of a reception algorithm and demodulation methods.

Consider first the methodology of noise immunity as applied to data transmission and then retrospectively review the milestones which some divisions of statistical communication theory have passed in their development.

I.3. Methodology of Noise Immunity

It is essential to draw a clear demarcation line between the reliability and noise immunity of a system. This is especially important because of the fast progress in microelectronics and integrated circuit technology.

There are two types of system failure to distinguish—irreversible failures and reversible failures. An irreversible failure is caused by such factors as ageing, corrosion, inadequate structural strength, etc. The evaluation of probability characteristics when estimating a non-failure operating time, given a possibility of irreversible failures, pertaines to reliability theory. Reliability can be improved by better technology, equipment redundancy, and some other means. A reversible failure is brought about by interference. In this case, the system recovers its serviceability as soon as the interference vanishes. Matters of optimal design leading to a noise-immune system relate to the theory of noise immunity.

Information transmission and reception involves both reliability and noise immunity. However, either may have an optimal solution different from that of the other. Let us consider the matter of noise immunity in more detail.

Often its objective is stated as filtering, that is separation of the desired information from the undesired. Accordingly, principal aspects are considered.

1. Separation of the desired information from interference without loss of the former. It is the fundamental objective of communication. If it cannot be achieved, for instance due to a limited bandwidth, then one usually seeks at least to minimize distortion so as to satisfy some optimality criterion.

2. Separation of a part of the desired information from the remaining useful information and interference, which entails the loss of some information. This is the fundamental objective of measurements. Here some distortion of a desired signal is not objectionable, such as when functionals of the primary information flow are measured. The information flow is reduced to functions and variables which describe the original process in some way.

There certainly may exist cases that do not fit the above classification or are intersection of the above two objectives.

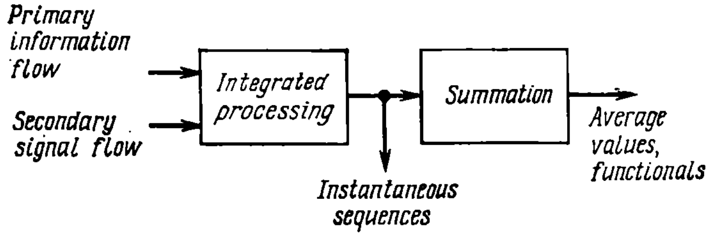

Fig. I.1. General block diagram of integrated system

The methods chosen to improve the signal-to-noise ratio materially depend on the type of signals and interference. It is convenient to treat as basic elements at least three types of signals and interference.

1. Signals differing in frequency. These signals occupy a relatively narrow frequency band and can be separated in terms of frequency. Examples are harmonic oscillations and narrowband processes.

2. Signals occurring at different times. These signals (periodic or aperiodic) appear at discrete moments of time, and their separation requires pulse techniques. Examples are pulse trains and discrete broadband signals.

3. Noise. It spreads over a wide range of frequencies and varies continuously with time. In theoretical studies, noise is supposed to have the normal distribution.

All the three types of signals can be used to model both information-bearing signals and interference. Accordingly, a total of nine combinations may be formed.

At the transmitting end, the choice of coding and/or modulation methods that would ensure noise immunity should be based on the a priori characteristics of the signals being transmitted and interference. At the receiving end, the choice of decoding or demodulation methods, in contrast, should be based on the a posteriori characteristics. The Bayes formula supplies the probability that the certain original message signal was sent if a signal corrupted by noise is received. In other words, one finds the probability of the cause from the probability of the effect in much the same way as we determine the input signal from the output signal and the known system function in the case of deterministic processes.

Figure I.1. shows a generalized block diagram of a data processing system which can equally apply to both the transmitting and the receiving end. Let us call it an integrating system. In this system, the primary information flow is combined with the secondary signal flow which may be deterministic or nondeterministic, for example, a pilot signal or a system function, or some other process with which the original process is compared. Now, the classification may be

carried out according to either secondary signals used or the union operations involved.

Integrated operations

multiplicative:
- convolution
- product
- polynomial product
- matrix product
- logical union

additive:
- algebraic addition
- modulo 2 addition

Processes being combined

deterministic:
- standard signals (Fourier or Walsh analysis)
- test signals (synchronous detector, matched filter)

nondeterministic
- delayed processes (auto- and cross-correlation)

We shall introduce now what we think is a convenient classification of methods designed to enhance noise immunity. Consider first the methods involving no reduction of information.

General methods: increasing the transmitter's power, spatial discrimination (using directional antennas), interference screening, spectral discrimination, time discrimination, compensation, and adaptation.

Special methods of data transmission: suitable techniques of modulation (FM, PM, PPM, PWM) and demodulation (in particular, PLL), or coding techniques (nonredundant coding, error-correcting codes).

Special methods of handling information: signal averaging, interference averaging.

The choice of a method is strongly affected by the a priori characteristics of interference. Thus, FM may prove ineffective in the presence of heavy interference because the power of the interference increases with increasing bandwidth. On the other hand, with PLL used at the receiving end, a narrower frequency band may be used, so the a posteriori characteristics must also be considered.

Under the conditions of weak interference, redundant coding is ineffective because binary signals can be regenerated. The redundant coding is worth using if the expected noise is either impulsive or is not normally distributed, or when rare fluctuations of interference might lead to an impermissible loss of information, as in data transmission.

The methods entailing loss of information belong to the measurement field. They serve to turn the original information flow into the measured values and may be divided in two groups, according to what is analyzed—the signals or noise. The first group is based on a comparison of the incoming signals with a deterministic reference signal (using the Fourier, Walsh, or Laguerre polynomial expansion in terms of the reference signal) or a nondeterministic signal (using auto- and cross-correlation analysis). The second group employs extrapolation of noise (autocorrelation, Kalman filtering) and suppression of multiplicative interference (e.g., smothing of fadings).

Of course, any classification and comparison of data transmission and processing techniques must inevitably involve an optimal noise immunity criterion constructed with allowance for the acceptable loss of some information. Yet, the decisive factors are the nature of noise and the type of signal.

In giving an outline of what noise immunity was in the past and what it will be like in the future, we cannot help referring to optimality criteria and relevant constraints.

I.4. Theory of Coding and Decoding

Coding and decoding are the subject-matter of information theory taken in its narrow sense (the so-called Shannon's information theory). In his pioneering work, Shannon introduced the measure of information, source entropy and channel capacity, proved theorems which were to govern the development of source coding theory (data compression) designed to remove redundancy, and theory of noise immune coding. The original findings were later verified and elaborated upon by Shannon himself [479, 480] and many other investigators, notably the leading Soviet mathematicians Kolmogorov [141], Khinchin [227], Dobrushin [99], and Pinsker [211a]. In one of his publications [141] Kolmogorov gave a rigorous mathematical form to Shannon's findings and set forth a high-priority research program in the field of information theory to solve practical problems of communication. Another work by Kolmogorov [418], published in 1956, examines the channel transmission rate subject to a given criterion of reproduction fidelity. Speaking in Moscow at the session of the Popov Scientific Society for Radio Electronics and Communications in May 1965, Shannon dwelt on the lower bound estimates to error probability for a broad class of discrete memoryless channels. He put the work as a most important achievement of American scientists in the field of information theory at the time. Later, he published the results in a common paper with Gallager and Berlekamp [481].

There was a time when communication systems had no coders and decoders. This equipment, indeed, is not of vital necessity in many

cases where the transmission problem can be solved by suitable modulation-demodulation techniques. The situation, however, is entirely different for the communication systems which must meet the stringent requirements to transmission rate and fidelity. Here coding comes to help compressing the data and making it more immune to the harmful influence of noise.

The optimism that was initially brought about by the Shannon results gave its way to the disappointment when these results were found nothing but the theorems of existence incapable of leading a practising engineer to realizable codes which could meet the requirements of practice. The wanted practical recommendations appeared in the late fifties, and the following decade saw a lot of them.

Message source encoding (data compression) is used in special voice and picture transmission capabilities and in space communications. The source encoding was a success in data transmission by the Soviet space probes Mars-6 and Mars-7 [69].

Of fundamental importance to error-free coding theory (error correcting codes) has been the relation between the code length, code distance and transmission rate. Found by Varshamov [68] and Gilbert [374], the relation gives the ultimate correction capability of a linear code and is known in the literature as the Varshamov-Gilbert bound. A large number of codes close to this bound have been obtained by purely algebraic procedures. Such are, for instance, the Bose-Chaudhuri-Hoequenghem (BCH) codes which are given by the roots of polynomials over the Galois fields [327, 387]. Another class of linear error-correcting codes is due to Goppa [89]; these codes are set up by polynomials over a Galois field. The BCH codes are the only cyclic codes in the class. Almost all Goppa's codes are asymptotic to the Varshamov-Gilbert bound.

Though attractive, the error-correcting codes have a serious limitation on the practical side—the respective decoders become too sophisticated as in many cases the number of decoding operations increases exponentially with the code length. The simplest method to realize the threshold decoding has been devised by Massy [435] and developed upon by Kolesnik and Mironchikov [138]. From the viewpoint of coding and decoding equipment, the concatenated codes proposed by Forney [365], and generalized later by Blokh and Zyablov [20, 24], hold promise for implementation. Concatenated coding will be treated in Ch. 3 of the book.

Unlike the algebraic procedures of decoding based on the structure of coding alphabets, the sequential decoding of convolution codes, which has been suggested by Wozencraft [505], uses a probabilistic approach. This decoding technique is asymptotically optimal in Shannon's sense. Its average number of operations required to decode a single symbol is a linear function of the code length. Further devel-

opment has been given to the technique in the work of Fano [361] and especially in the research of Zigangirov [103, 104] who designed a maximum likelihood algorithm to sequentially decode convolutionary codes. This algorithm is discussed in more detail in Ch. 3 of the book.

Korzhik, Osmolovsky and Fink have suggested a new statistical approach to the coding in feedback systems. When used in binary channels, the coding makes the probability of unrevealed error not exceed a given value decided by the parameters of the code rather than the channel properties [146]. In the circumstances, if the quality of the channel worsens, the transmission rate becomes slower rather than erroneously received symbols appear.

In the last few years, general models of channels which make allowance for errors and defects have been developed and studied. The problem of defect correction arose due to the operation of data storage facilities: some memory cells with binary symbols may be damaged and read out one and the same symbol irrespective of what symbols are stored in them.

An informational concept of defect has been introduced in the work [155]. Tsybakov in his paper [282] has considered the joint correction of defects and errors. There has been formulated the necessary and sufficient condition to correct defects and errors by a given code. The upper bound has been obtained for the rate of a code which is able to correct a given number of defects and errors. Defect correction in informational sense is deeply involved into the problem of data transmission over broadcast channels.

I.5. Reception of Signals

Noise immunity analysis of reception circuits. Solving of the analysis problems was one of the first contributions to statistical radio engineering and statistical communication theory. In the 1940s, Rice [456], Bunimovich and Leontovich [36-38] obtained the expressions for correlation functions, power spectra, an average number of axis crossings and distribution functions of a deterministic signal in additive Gaussian noise after transformation in the typical elements of reception circuits (amplifiers, detectors, limiters and filters). These findings were soon used to effectively analyze noise immunity of various systems. Thus, already in 1946, Shchukin in his work [300] gave an analysis of the known system: the broadband clipping amplifier—matched filter, and Siforov discussed a probabilistic approach to the influence of interference on the reception of impulse signals in the paper [244]. A large number of research papers dealt with the design of the noise-immunity characteristics for multichannel radio-relay links using both pulse and frequency modulation techniques. Suffice it to note two papers by Borodich [33, 34]

which were among the first works on the subject. The probability analysis of the phase of narrowband random processes [161, 281] proved to be a successive tool in evaluation of noise immunity for the systems employing phase-difference-shift keying (PDSK) devised by Petrovich [209]. Another illustrative proof that the probability methods have a productive potential was supplied by the research on the error-free reception of signals corrupted by multiplicative noise in fading channels (see also Sec. 7.3). The theory of pulse random processes [142, 161] has played a significant role in analysis of noise immunity for communication systems with pulse modulation and TDM channels and those with asynchronous addressing.

Optimum filtering. In the above works of Kolmogorov [140] and Wiener [506], the problems were set up and solved of how to extract (estimate) an unobserved random signal from additive random noise by a linear filter optimally in mean square error. The characteristic of an optimal filter can be found by the Wiener-Hopf equation. In studying this equation, Krein, a Soviet mathematician, derived outstanding results [153]. The theory of optimal linear filtering is indebted to many authors in that its field of application has been materially widened to include the physical realization of filters, the aspects of finite observation time, and process nonstationarity. To mention only the recurrent algorithm used in filtering of nonstationary Markovian processes, which has been advanced independently by Stratonovich [251] and Kalman [400] and found important practical applications. In the literature this algorithm is known as the Kalman filter [467]. A discrete version of the algorithm is successfully employed in data transmission systems, notably in space communications [318, 429]. An exhaustive bibliography on the theory of optimal linear filtering (390 entries) is to be found in the review compiled by Kailath [399].

In synthesizing optimal demodulators for FM and PM systems, use is made of nonlinear random process filtering. Initiated by works of Stratonovich [250, 251], the Markov model of receiving signals has made more headway. A statistical approach was used to synthesize optimal, linear and nonlinear phase-lock loops which were employed as synchronizing means and as synchronous phase discriminators [261a, 295, 496].

Optimum detection of signals. Signal detection, the major task of communication, was formulated in the doctor thesis by Kotel'nikov [151] as a multiple-alternative test of statistical hypotheses; Kotel'nikov also obtained an algorithm of optimal detection subject to a maximum posterior probability criterion. The following decade saw an explosion of activity in the optimal statistical synthesis of information systems. More complicated models of signals and interference and more general quality criterion were considered. The period culminated in what became later the classical theory of opti-

mal statistical synthesis, which with a various degree of completeness was treated in almost simultaneously published monographs by Middleton [441], Vainshtein and Zubakov [44], Helstrom [386], Gutkin [95], and Fink [267]. Helstrom discussed narrowband random processes in his book, by applying in an elegant style the complex representation that had been introduced by Gabor [370].

Presently, the theory of optimal statistical synthesis is based on the Bayes minimum average risk criterion. The expression of Bayesian averaged risk is a complicated functional of random process applied to the receiver input. The Bayesian algorithms of decision making and estimation of unknown signal parameters are obtained in a closed form and may be relatively easily realized by technical means in the case when the observed process is an additive mixture of a signal and normal noise. In the circumstances, the linear section of the optimum receiver is a set of matched filters. As should be recalled, the impulse response of a matched filter exactly coincides with the signal as reversed and shifted in time. This result is due to Narth [447] who derived it in 1943 when seeking for a linear filter that would maximize the ratio of squared peak intensity of signal to noise variance. Somewhat later it became known that the matched filter was also of correlator variety.

Although an optimum system, as predicted by statistical synthesis theory, is not always feasible technically or sound economically, its probability analysis is, nevertheless, worthwhile to evaluate how short of it a physically realizable, though suboptimum, system falls. In recent years jointly optimal Bayesian algorithms of signal detection and parameters estimation (filtering) have been obtained [178].

Overcoming the a priori uncertainty constraint. The lack of a priori data often prevents the results of Bayesian optimal synthesis from being put into practice. Even if some assumptions are introduced, the solution obtained under one assumption may prove to be substantially nonoptimal under another. To overcome prior uncertainty, the theory of optimal reception is being developed towards the systems which are close to optimal ones, do not change the desired parameters when signals and noise vary, and lend themselves to practical implementation [164, 208, 226, 268]. Use is also made of adaptation, nonparametric statistics, and appropriate approximations (see also Sec. 5.1).

I.6. Theory of Signals

Signal synthesis is a major objective in communication system design. As the joint optimization of transmission and reception is still an unsolved problem, the procedure in partial optimization is as follows. First an optimal reception algorithm is evaluated for an

unspecified, though known, form of signal, then such a signal is selected to this algorithm that will minimize the probability of error in the class of possible signals. Thus, for example, opposite polarity signals give the best error probability performance when extracting signals in additive normal white noise in binary FSK, PDSK, and on-off keying systems.

Matched filtering utilized as the key element of optimal reception has stimulated both the theoretical synthesis of signals which are well auto- and cross-correlated and pulse compression technique. The initiating work in the field was the book by Woodward [504] published in 1953, who suggested the uncertainty function now in wide use in signal synthesis theory. The following years saw a new class of signals having a large time-bandwidth product. The signals of the class are called differently: wideband signals, composite signals and pseudonoise signals. Henceforth we will use the last of the terms (see Chap. 4).

That a pseudonoise signal (PNS) is compressed when passing through a matched filter was noted as early as 1956 by Shirman [298], and in 1960 a series of publications on the subject followed [16, 343, 415, 491]. One of the first PNS which gained wide application was the linear frequency-modulated pulse train [343]. Along with analog-modulated signals, PNSs based on discrete codes appeared to influence the parameters of carrier oscillation. These PNSs may be divided into three groups. The first includes the pulse trains having a constant carrier frequency [460]. A large number of practically important PNSs makes up the second group of PSK signals which are based on the Barker codes [317], recurrent sequences of maximum length (M-sequences) [390, 510], and polyphase codes [275, 367, 375, 376]. The third group embraces discrete frequency sequences [344]. In the Soviet Union, the general methods of PNS synthesis are due to Vakman [37], Varakin [51], Sverdlik [238] (see also [48]). The first practical radio communications system where PNSs were used to advantage to eliminate crosstalk in multipath channels apparently was the Rake system [455]. The following years saw a number of PNS applications in communications. To note are the space communication systems Mariner [313], Digilock [310], and communication satellite systems [314].

I.7. Statistical Multiplexing of Signals

In multichannel communications systems the subscriber signals are joined or multiplexed into a single composite signal. Effectiveness of multichannel communications can be materially improved by taking into account the statistical structure of the message source. The structure is characterized, firstly, by the distribution functions of instantaneous values of the transmitted message and, secondly, by

activity, i.e. the probability that the source deemed to be operating does produce a message. This activity concept may be well illustrated by an example of a speech message in telephone conversation. In any conversation there are spacings between syllables, words and sentences, and, besides, each subscriber speaks only a part of conversation time, since he also listens to the party.

The first equipment making allowance for the activity statistics was installed on the frequency-division multiplexed transatlantic cable link (London-New York) in 1960 and became known as the Time Assignment Speech Interpolation (TASI) [331]. Speech messages were transmitted in portions decided by the intervals of subscribers activity, thus raising the trunk capacity twice.

In 1960 another modulation technique was suggested named by its authors as interval pulse-position modulation. This technique utilizes both activity statistics and instantaneous message value statistics in time-division multiplexed channels [165, 166]. One more statistical time-division multiplex was developed by Ach [10] in 1964 (see also Secs. 6.3 and 8.4 below).

1.8. Communications Networks

In designing multi-access communication systems, the principles of statistical theory are also employed. One of the first system of the type was the Random Access Discrete Address network (RADA) [355] based on asynchronous addressing multiplex. Another example of how the principles of statistical multiplex can be implemented is the ALOHA system [305, 306] designed to communicate the central processor and the user terminals involved in the exchange system. A random-access radio channel has been developed at the Budapest Technical University for a data transmission system (see Sec. 8.4).

Multi-access principle is the basis for designing the modern satellite communication systems. A classification of satellite communication systems operating on the multi-access basis is given in [72].

The trend in communication system development is that in the near future the communication networks will mainly carry digital data between computers. A network of future will be a complex multi-access system linking a great number of computers with a central processor available to a multitude of terminals. State of the art and trends in the development of complex systems combining computers and information-carrying channels are treated in the book by Samoilenko [234]. Some actual problems of the trend in communication engineering will be discussed in the two last chapters of the book.

TABLE I.1

15-year period	Year	Work, event or development
Theoretical prehistory	1928	Works on general (deterministic) communication theory [382, 448]
	1930	Wiener's work on generalized harmonic analysis [499]
	1934	Khinchin's work on the correlation theory of stationary stochastic processes [406]
	1939	Kolmogorov's work on random process filtering [417]
	1942	Wiener's work on random process filtering [500]
	1943	Narth's matched filter [447]
	1944	Rice's work "Mathematical analysis of random noise" [456]
	1946	Works of Shchukin, Siforov, Leontovich, and Bunimovich [300, 244, 37]
Period of flourish (theory and practice)	1947	Doctor thesis by Kotel'nikov [151]
	1948	Shennon's work "A mathematical theory of communication" [478]
	1953	Advance of the uncertainty function by Woodward [504]
	1956	Compressión of pseudonoise signals in a matched filter [298]
	1957	First Soviet monographs on stochastic process theory in radio engineering and automatic control [161, 216]
	1957	Advance of sequential decoding [505]
	1959	Employment of M-sequence [510]
	1960	Advance of recurrent Kalman filtering algorithm [251, 400]
	1960	Completion of the Bayes theory for receiver synthesis [441]
Contemporary developments (theory and practice)	60 s and 70 s	Satellite communication systems and long-range spaceprobe communication systems High bit-rate digital communication systems, PNS and multiplex techniques Multiaccess systems Optical communication systems Computer-communication networks Large-scale integration circuits

I.9. The Chronology of Statistical Communication Theory

The brief chronological table (Table I.1) illustrates to some extent the dynamics of development of the statistical communication theory and its applications.

I.10. Literature, Conferences, Symposia

As the number of research papers and specialists who use statistics and probability methods in tackling the problems of communication, radio engineering and automatic control increased, a demand arose for books that would generalize and systematize the theoretical results. First publications in applied theory of random processes appeared in the Soviet Union in 1957 [161, 216], then in the United States [323, 354]. An important part in popularizing the statistical communication theory was played by the books of Kharkevich, a Soviet academician [272, 273]. The subsequent years saw, now well known, monographs [28, 45, 109, 132, 136, 267]. Of the books published outside the USSR in this period, the lectures delivered by the American scientists at the Massachusetts Institute of Technology in 1959 [425] and at the University of California in 1963 [315], publications [362, 453, 506] and books by Prof. Lange of GDR [421, 422] are to be noted.

In 1974, Svyaz Publishers, Moscow, opened a series under a common heading "Statistical Communication Theory" to mirror the state of the art and tendencies in the different branches of the theory. Among the books of this series are [24, 104, 147, 296, 467, 473, 486].

Of great importance for the development of statistical communication theory are the scientific conferences, meetings and symposia organized in the Soviet Union by the Popov Society for Radio Electronics and Communication and by the Cybernetics Council at the USSR Academy of Sciences. The international symposia on information theory regularly held in the Soviet Union (Dubna-69, Tsakhkadzor-71, Tallin-73, Repino-76, Tbilisi-79) became very popular. The All-Union conferences on coding and data transmission theory also promote the development of information systems. The materials of the 1970 All-Union conference on optimal statistical synthesis of information systems under prior uncertainty made up a subject publication of "Radioteknika" journal [167]. It is only to add that this book is based on the materials delivered by the scientists from the socialist countries at the conference on statistical theory of communication and its applications held in Pushchino, near Moscow, in 1977.

I.11. Conclusions

The concepts of probability theory and statistics have proved in the last decades to be very fruitful for all branches of radio engineering both theoretically and practically. However, as sometimes it is the case, with a new methodology extreme points of view appear. There are examples of stale application of the deterministic principles in situations where they are ineffective. The traditional sinusoid is forced to do a job which can be easily fulfilled by a stochastic process. On the other hand, not infrequently the probabilistic methods are applied in an unqualified manner just for novelty.

The significance of statistical communication theory has risen drastically after the appearance and progress in computer and large scale integrated circuit technology, which enables one to materialize many theoretical concepts of coding and decoding, transmission and reception, rejected earlier as being too abstract and nonrealizable. The achievements and possibilities of the theory are one of the major factors for scientific and technical progress.

Chapter

1

Stochastic Models of Signals, Noise and Communication Channels

1.1. Basic Principles of Probabilistic Modeling

Introductory remarks. The term 'communication channel' is used to denote a certain part of a communication system which may vary in communication equipment involved, depending on a particular task. So, for instance, in dealing with encoded messages (coding theory) the channel is the whole set of facilities linking the coder and decoder. In handling continuous signals (signal transmission theory), the channel embraces a lesser number of physical units because many elements of both the transmitter and receiver (e.g., a signal conditioner, modulator, demodulator, etc.) the designer does not regard as giving "fixed" processes and therefore puts out of the channel.

The channel is also understood as the medium carrying the signals. For instance, Shannon [478] has termed the channel as the medium used to convey signals from the transmitter to the receiver; thus, the channel may be a pair of wires, a coaxial cable, space where radio waves propagate, a beam of light, etc.

Communication channels are generally classified according to the nature of their inputs and outputs. If the input to a channel is discrete and the output is also discrete (coding theory), the channel is said to be *discrete.* If the input and output are both continuous, the channel is said to be *continuous.* If the input is discrete and the output continuous, the channel is said to be discrete-to-continuous.

An approach to describe signals and interfering noise depends on the channel in question. Figure 1.1 depicts frequently used block diagrams of systems transmitting discrete and analog messages. The models discussed below are those most often used to represent a process $X(t)$ at the output of a continuous channel. We shall call it the observed process. Statistical models of the process $X(t)$ are used to optimize the structure of receivers, decision making equipment or demodulators and to analyze the performance of receivers of fixed structure in the presence of noise.

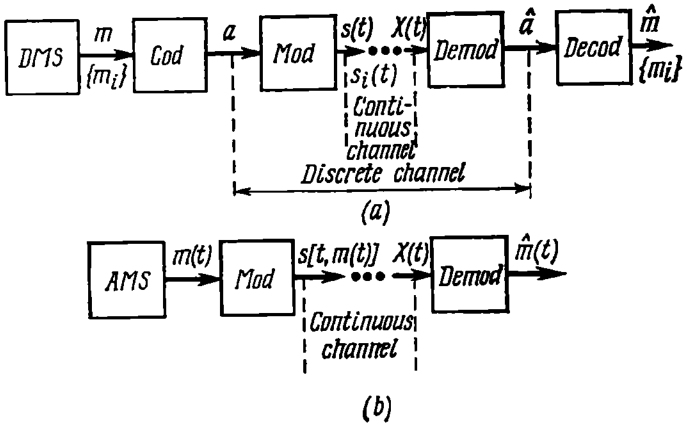

Fig. 1.1. Block diagram of communication system for (*a*) discrete and (*b*) analog message source: DMS—discrete message source; AMS-analog message source; Cod—coder, Decod—decoder; Mod—modulator; Demod—demodulator

As a rule, the process $X(t)$ is a mixture of a desired signal $s(t)$ and interfering noise $\xi(t)$; the characteristics of this process are entirely different in different channels and communication systems. Though straightforward, it would be unwise to devise a new mathematical model for every new system. A possible way is to suggest a mathematical model which would sufficiently describe the process $X(t)$ so that any problem might be solved for any communication system. However, such a model would be cumbersome and for many problems the solutions produced would hardly be of value. Hence, it is desirable to have simple models which would be adequate to at least a narrow class of real processes and would generate closed solutions when dealing with problems of optimal synthesis and analysis.

Noise. Noise may be defined as disturbance (electromagnetic, electric, acoustic, etc.) impeding the reception of a desired signal [274]. It can unpredictably change the shape of signal and information it carries. The distortion of a signal due to noise should be distinguished from that appearing when the signal passes through various elements of radio equipment. In the latter case, the distortion introduced by each piece of equipment may be related to the signal by a determinate operator and thus compensated, if need be.

Depending on the origin or the type of source, interfering effects may be classed into extraterrestrial, atmospheric, man-made, caused by the channel (echo, fading, fluctuating path attenuation, etc.); jamming; thermal noise, etc. According to the manner in which noise acts on a signal it is distinguished into additive and multiplicative. Additive noise often results from emission of external sources,

crosstalk due to the finite amount of protection between the pair of cables, interference due to the nonlinear response of the circuitry; thermal noise; discrete characteristics of electromagnetic radiation or light displayed at a macroscopic level when the signal is weak, etc. Typical examples of multiplicative interference are unpredictable changes in channel gain, interruptions, etc.

Owing to the fact that momentary values of interference are unpredictable, they are often treated as stochastic processes of various probability characteristics, as a rule other than those of information-bearing signal. There exists a general method to statistically describe a random process [79], which may be applied to specify any interfering effect in communication channels. This method envisages that a probability measure is stated on the sets of realizations of these interfering effects. However, to describe the sets and specify probabilities on them is not an easy matter. This would require a sophisticated mathematical formulae of the measure theory on a functional space to be applied instead of simple traditional techniques used in the communication engineering.

The probabilistic description of noise would be drastically simplified if the classes of interfering effects are appropriately selected. Then, the noise of each class may be defined in a different way choosing the most economical approach which gains closed solutions for many problems of statistical communication theory. From this viewpoint it seems practically sound to divide noise in communication channels into the following classes: impulse, fluctuation, and quasi-deterministic.

Fluctuation noise is most frequently met. This class usually involves thermal noise, nonlinear and linear transient influence and echo-signals in multichannel relay and cable lines, many kinds of electromagnetic or audio frequency emissions in radio and hydroacoustic channels. The fluctuation noise is most often caused by random deviation of physical quantities from their mean values. Fluctuations, in turn, are due to discrete structure of matter and statistical nature of the majority of physical phenomena and quantities. For instance, random deviations of current in an electric circuit from its mean value, even at the macroscopic level, result from the discrete nature of charge-carriers (electrons and ions), and random deviations of potential difference from zero value in any conductor may be due to the thermal motion of charge-carriers in this conductor. These fluctuations cannot be practically eliminated. Another example of uneliminated fluctuation noise is electromagnetic radiation which is so due to the radiation's discrete nature.

Impulse noise is also widely met. It can be exemplified by the arrivals of overload intervals in transmission paths, short interruptions (dropouts) and routine switching in the link and supplies, lightnings, etc. This type of noise is a major factor limiting the fidelity of

reception. About one fifth of all errors in binary channels is due to impulse noise. Short breaks are the cause of almost 80% of errors in binary data transmission systems [30, 126].

Quasi-deterministic noise is often a significant contributor to the total power of interference. This class may include, among others, the narrow-band signals of external sources, unmodulated portions of carriers, pilot and supervisory frequencies leaking into the channels allocated bandwidth.

The above noise classification is the most suitable for problems of statistical communication theory since the stochastic models are thereby appropriately simplified, the classes of noise being comparatively broad. Applying the classification, constructive solutions for the majority of the problems can be sought, such as the performance analysis of noise-limited reception, and synthesis of receiver circuits based on some probabilistic optimality criteria.

Stochastic modeling of fluctuation noise will be treated in detail below.

Signals. Of all signals encountered in a communication system, as desired are qualified those which carry useful information. The signals which are precisely known prior to reception (deterministic) cannot serve as information carriers [1]. Though the ignorance of a message sent by a source can be presented in a variety of ways, we will assume below that the ignorance displays itself as an inability to precisely predict the values of a signal in future even though all the values of the signals in the past up to a given moment of time are known. This assumption is generally accepted and seems to be intuitively sound. If we now assume that a hypothetical ensemble of signals exists and features a statistical stability, then the signals become possible to treat as random processes. Hence, if the above assumptions hold, then the signals should be described in the same way as noise, that is via some probability models. Therefore the methodology of modeling the signals and noise in each particular problem of communication would be the same, so they could be treated on a common basis. The construction of the stochastic models will be evaluated using as an example a real process and the results of the modeling will be discussed on the final stage.

Working towards this probabilistic presentation, it would be helpful to class the signals analogously to how it has been done for noise. The classification may look as follows:

— fluctuation signals (e.g., music, voice, television broadcasting signals, composite signals of multichannel communication);

— impulse signals (e.g., arriva ls of calls to secure a circuit or a toll center, signals in multichannel asynchronous data trans-

[1] This does not exclude from consideration the detection and extraction of deterministic signals from noise, because these processes involve uncertainty in respect to certain events related to the signals.

mission systems, an arrival of signals at the output of multipath channel);

— quasi-deterministic signals (e.g., single sideband carriers, surveilance frequencies, intelligence signals in digital data transmission).

The general stochastic model. Consider now the general model that could be employed to mathematically describe signals and noise belonging to different classes. On the advantage side of this model is its all-embracing character. On the disadvantage side are the difficulties that the designer meets trying to get the required by the model data on the real process, then trying to evaluate its goodness of fit to the real process, and, finally, it proves too unwieldy to yield analitical solutions of many communication problems.

Yet, we will treat the model to illustrate how the unified methodology can be applied for constructing or choosing the stochastic models to fit real processes, and what could be sacrificed to simplify the model. We will enumerate the major requirements posed on the probability models in communication. Besides, we should point out the existence of general type processes for which this model, though cumbersome, is the only feasible.

Consider a hypothetical experiment E, one trial of which is an observation of the values the observed process $X(t)$ has over the time interval $[0, T]$ [1]. Let every trial be made subject to one and the same set of conditions $\mathscr{Y}$ and the outcome of a trial be not influenced by the prior outcomes and, in turn, give no effect on the outcomes of future trials. The outcome of every observation in this experiment is a function of time over the observation interval $[0, T]$. And the collection of all outcomes will thus be the collection of functions over the interval $[0, T]$, related to the given experiment E. We will denote this collection of functions by the symbol $\mathscr{X}_{[0, T]}$. To emphasize that the functions, as a rule, differ from trial to trial, we denote the function of the jth trial $x^{(j)}(t)$, $0 \leqslant t \leqslant T$, $j = \overline{1, N}$, where N is the number of trials in a given series of measurements.

If we deal with a random experiment E having a statistical stability [2], and if the outcome of each observation is a function over the interval $[0, T]$, then we may say that in the experiment E we observe the random process $X(t)$, $0 \leqslant t \leqslant T$. The functions $x^{(j)}(t)$, $0 \leqslant t \leqslant T$, $j = 1, 2, \ldots$, are referred to as the realizations of the process, and the collection $\mathscr{X}_{[0, T]}$ of all the realizations is called the

---

[1] For example, in PCM systems this interval $[0, T]$ may coincide with the length of an elementary symbol or code word composed of several such symbols; in analog communication systems, this may be an arbitrary interval, including the infinite one.

[2] A manner in which the statistical stability manifests itself in this type of random experiment will be discussed later.

ensemble of realizations or the sample space of the process $X(t)$, $0 \leqslant t \leqslant T$.

Examples of realization ensembles.

Example 1. Let a process $S(t)$ be the set of cosinusoids

$$S(t) = a\cos(2\pi f_0 t + \Phi),\ 0 \leqslant t \leqslant T \tag{1.1}$$

where a and f_0 are the known parameters of the process, and Φ is a continuous random variable assuming various values of the interval $[-\pi, \pi]$ in various observations. The sample space of the process is thus

$$\mathscr{X}_{[0,T]} = \{a\cos(2\pi f_0 t + \varphi),\ 0 \leqslant t \leqslant T:\ -\pi \leqslant \varphi \leqslant \pi\} \tag{1.2}$$

that is, it is an uncountable set.

Example 2. Assume that the process being observed is the additive mixture of the process (1.1) and fluctuation noise $\xi(t)$

$$X(t) = S(t) + \xi(t),\ 0 \leqslant t \leqslant T \tag{1.3}$$

and the realizations of the process $X(t)$ can be observed at the output of a real channel having a limited bandwidth. The latter constraint makes so that all the realizations of the process (1.3) are continuous functions over the interval $[0, T]$. The sample space of the process $X(t)$, $0 \leqslant t \leqslant T$, may now coincide with the space $C^1_{[0,T]}$ of all functions continuous over the interval $[0, T]$. This space, in particular, includes as a subset the sample space (1.2).

From the above examples it is seen that various processes may have various realization ensembles. Therefore, first thing we should do designing the stochastic model of a process is to specify the respective sample space. For example, in statistical communication theory, the stochastic models of signals and noise are devised as the processes such that the sample spaces of both the mixture of signal and noise and noise alone should coincide. Otherwise, the model would produce singularity, that is a possibility to completely reject the noise when observing over arbitrarily small time intervals, which, certainly, would contradict to what we have in reality.

The sample space is not deemed an important characteristic of a process—in applied works it is not even mentioned in description of a model used. However, methodologically, this notion proves very important and often leads to sufficient simplifications in theoretical structures involving the stochastic models.

Examples of events related to the observed process. The main objective of the probability model for any process observed is to calculate the probabilities of events related to the process. Examples of such events are given below.

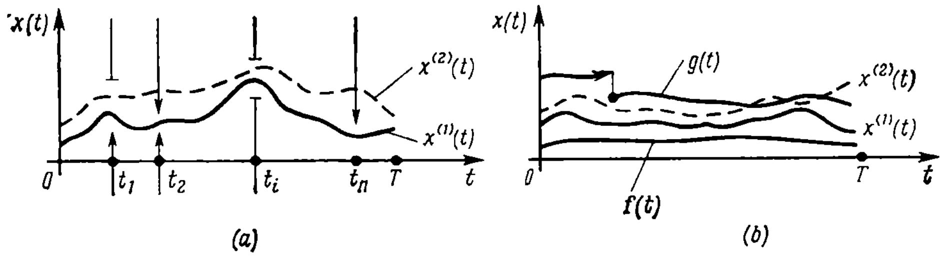

Fig. 1.2. Examples of events related to process being observed

Example 3. An event A_n related to process $X(t)$, $0 \leqslant t \leqslant T$, will be defined as follows:

$$A_n = \{X(t_i) \in I_i,\ i = \overline{1, n}\},$$
$$I_i \in R^1,\ i = \overline{1, n} \tag{1.4}$$

The event A_n takes place on the jth trial of experiment E if the realization $x^{(j)}(t)$, $0 \leqslant t \leqslant T$, of the process occurs such that all the following conditions are satisfied: $x^{(j)}(t_i) \in I_i$, $i = \overline{1, n}$. If at least one of the above conditions does not hold, then what is observed is called an opposite event $\bar{A}_n$. In Fig. 1.2*a*, the sets I_i, $i = \overline{1, n}$, are depicted as the "gates" (intervals) through which the realization $x^{(1)}(t)$, $0 \leqslant t \leqslant T$, is to pass for the event A_n to occur. If the realization $x^{(2)}(t)$, $0 \leqslant t \leqslant T$, occurs, then we shall observe the event $\bar{A}_n$.

In a particular case of $n = 1$, $t_1 = T$, $I_1 = \{x : x \geqslant h\}$ we obtain the event $A_1 = \{X(T) \geqslant h\}$, the occurrence of which is often interpreted in statistical communication theory as detection of the desired signal in a noise background.

Example 4. An event A_T related to the random process $X(t)$, $0 \leqslant t \leqslant T$, is defined as follows: $A_T = \{f(t) < X(t) \leqslant g(t),\ 0 \leqslant t \leqslant T\}$, where $f(t) < g(t)$ and both $f(\cdot)$ and $g(\cdot)$ are some deterministic functions. The event A_T will take place on the jth trial, if such a realization $x^{(j)}(t)$, $0 \leqslant t \leqslant T$, occurs for which the inequalities $f(t) < x^{(j)}(t) \leqslant g(t)$, $0 \leqslant t \leqslant T$, hold true. If at least one of the inequalities is not valid for at least one value t of the interval $[0, T]$, then the event $\bar{A}_T$ occurs (see Fig. 1.2*b*).

In a particular case, when $g(t) \equiv h$, $f(t) \equiv -\infty$, $0 \leqslant t \leqslant T$, the event is $\bar{A}_T = \{\max\limits_{0 \leqslant t \leqslant T} X(t) > h\}$. In statistical communication theory, when the event of the type occurs, it may be interpreted, say, as a channel overload on the interval $[0, T]$.

Thus, the probability model of a process should, in addition to the sample space $\mathcal{X}_{[0,\ T]}$, provide a possibility to calculate the probabilities of occurrence for the events of the A_n and A_T type without recource

to any trials. A problem, however, may require prior probabilities to be obtained for a variety of events related to the process. It implies that the probability model of a stochastic process should, in general, lend itself to computing the probabilities of any events related to the process. This is the second and most important requirement to the model of general type.

To sum up, the term "general stochastic model of the observed process" as applied to the theoretical and practical problems of communication will involve all the information available on the process (functions, tables, diagrams, etc.), which is used to compute the probabilities of all the desired events related to the process.

Stochastic model for the events of the A_n type. We shall now demonstrate which data will be sufficient to compute the probabilities of A_1 type of events [see Eq. (1.4) at $n = 1$]. It is easy to see that to decide the occurrence of the event A_1 on the jth trial, it will be sufficient to observe the value of realization at the time t_1 rather than all the realization $x^{(j)}(t)$ of the process over the interval $[0, T]$. In the circumstances, the outcome of every trial will be a single number. Hence, in the random experiment E, what is observed is a random variable $X(t_1)$ which for the sake of convenience we will henceforth denote X_1. Thus, the problem of seeking the stochastic model of a random process $X(t)$, $0 \leqslant t \leqslant T$, to compute the a priori probabilities of type A_1, has reduced to choosing a model of scalar random variable X_1.

What is that minimum amount of knowledge on a random variable X_1 which should be at hand to compute the probabilities of events related to this variable? The answer would, certainly, depend on which events are suggested to evaluate. If the class of the events is too broad and coincide, say, with the class of all possible events [1], then the stochastic model emerges to be too complicated. Therefore we have to abandon our efforts toward the stochastic model of random variable, that calculates the probabilities of *any* event related to the variable. If, on the other hand, we bound ourselves to a simple model which could give us access to the probabilities of only few events, the resulted probabilistic pattern of the variable would be impracticably coarse.

A tradeoff solution has been found in probability theory owing to a limitation imposed on the class of events under consideration. Accordingly, it is suggested to choose the stochastic models of random variables sufficient to compute probabilities of events belonging to the class which is more narrow than the class of all possible events, but still representative enough to involve *practically all events* of interest to applications. This narrow class is called a Borel set and

---

[1] The class of all possible events in this case may be described as the class of all possible subsets in R^1.

denoted by $\mathscr{B}_1$ [1]. From a course in probability theory (see, for example [139]) it may be recalled, that if for a random variable $X_1 = X(t_1)$ the function $F(x_1, t_1) \triangleq P\{X(t_1) \leqslant x_1\}$, $-\infty < x_1 < +\infty$, is given, then using this function, probability of any event belonging to $\mathscr{B}_1$ and related to $X(t_1)$ can be computed. The function $F(x, t)$ is called a *probability distribution function of a random variable* $X(t)$ or one-dimensional distribution function of a process $X(t)$, $0 \leqslant t \leqslant T$. Knowing the probability distribution, we have at hand that minimum amount of information about the random process that is required to solve the problem concerned. Hence, the one-dimensional distribution function is a sufficient stochastic model in this problem. Solving these problems, different functions may be interpreted as different stochastic models of a process.

The above argument leads to the following expanded statement: to compute the probabilities of events related to the values of a process $X(t)$, $0 \leqslant t \leqslant T$, for any arbitrary set of n t-values $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T$ belonging to $\mathscr{B}_n$, it is sufficient to know the n-dimensional joint distribution function of the process[2], that is $F(x_1^n, t_1^n) \triangleq P\{X(t_i) \leqslant x_i,\ i = \overline{1, n}\}$, $-\infty < x_i < +\infty$, $i = \overline{1, n}$ [139]. Here, $\mathscr{B}_n$ is the class of all Borelian events including *practically all* the events A_n^3 treated in communication theory. Thus, the nth order distribution $F(x_1^n, t_1^n)$ is the minimum necessary stochastic model of a process $X(t)$, $0 \leqslant t \leqslant T$, in problems where probabilities of events A_n are to be computed. These computations will be materially simplified if instead of $F(x_1^n, t_1^n)$ use is made of the joint probability density function defined as

$$w(x_1^n, t_1^n) \triangleq \frac{\partial^n F(x_1^n, t_1^n)}{\partial x_1\, \partial x_2 \ldots \partial x_n}$$

Accordingly, the probability density function $w(x_1^n, t_1^n)$ may also be viewed as a stochastic model of the process, sufficient to obtain probabilities for events A_n belonging to $\mathscr{B}_n$.

To sum up, in problems where it is required to compute the probabilities of events related to the values of the process at the finite number of time points $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T$ in the observation interval $[0, T]$, the stochastic model adopted should be either the n-dimensional distribution function $F(x_1^n, t_1^n)$ or the nth order probability density function $w(x_1^n, t_1^n)$ of the process. These functions completely describe the process in the problems of the above type.

---

[1] The class of Borelian events may be described as a class of sets in R^1 which can be obtained from the intervals of the type (a, b), (a, b), $[a, b]$ by repeated applications of operations of union and intersection to these intervals.

[2] From now on $(x_1^n)' = (x_1, x_2, \ldots, x_n)$, $(t_1^n)' = (t_1, t_2, \ldots, t_n)$.

[3] The class $\mathscr{B}_n$ is described as the class of Borel sets in n-dimensional Euclidean space R^n.

Stochastic model for the events of the A_T type. What amount of information on the process $X(t)$, $0 \leqslant t \leqslant T$, should be available prior to the trials this time to compute the probabilities of events A_T (see Example 4)? Answering this question without posing some additional restrictions on the process concerned would be a much harder task. This is so because this time to establish the fact that A_T occurs in the jth trial we would have to record all the values of realization $x^{(j)}(t)$ at all the times t of the continuous interval $[0, T]$. Thus, the set of values to be observed is no longer finite, or even countable, but, instead, is a continuum.

First, we shall demonstrate what is necessary to know about the process $X(t)$, $[0, T]$, to obtain the probabilities for events of the type $A_\infty = \{f(t_i) < X(t_i) \leqslant g(t_i),\ i = 1, 2, \ldots,\}$, where $\{t_i, i = 1, 2, \ldots\}$ is the set of rational numbers in the interval $[0, T]$. The event A_∞ is simpler than A_T in that it is defined by the values of realizations on only a denumerable set of points of the interval $[0, T]$. However, it is more complex than any event A_n as now there would be insufficient to know a finite-dimensional distribution of the process for arbitrary large yet finite set of rational time points $t_{l1}, t_{l2}, \ldots, t_{ln}$. According to the Kolmogorov theorem [139], to compute the probabilities of events A_∞ related to a random process $X(t)$, $0 \leqslant t \leqslant T$, the family of finite-dimensional distributions $\{F(x_1^n, t_1^n),\ n = 1, 2, \ldots\}$[1] for the process should be given. It implies that the family of finite-dimensional distributions may be treated as a stochastic model of the process $X(t)$, $0 \leqslant t \leqslant T$, sufficient to yield the probabilities for this type of events. It is to notice here that practical computation of the probabilities might be a cumbersome task.

It is important to mention the following fact: practically all events of any interest to communication theory and applications and related to the values of a process $X(t)$, $0 \leqslant t \leqslant T$, in rational points $\{t_i, i = 1, 2, \ldots\}$ on the interval $[0, T]$ belong to $\mathscr{B}_\infty$[2]. To conclude, we should note that the set of all rational numbers of an interval $[0, T]$ is not an exclusion in the above sense and that the family of finite-dimensional distributions, as defined on any countable set $\mathscr{T}$ of $[0, T]$, enables the probabilities to be computed for the events of the

---

[1] Let $t_1, t_2, \ldots, t_n$ be arbitrary values of an argument t in an interval $[0, T]$, and the respective random variables $X(t_i)$, $i = \overline{1, n}$, have a joint distribution function $F(x_1^n, t_1^n)$. Then the family of all these distributions for $n = 1, 2, \ldots$ and all pertinent t_i is called the family of finite-dimensional distributions for the process $X(t)$. So, to compute the probabilities of events A_∞ it will suffice to specify the family of distributions over the set of rational points of the interval $[0, T]$.

[2] The class $\mathscr{B}_\infty$ of Borel sets is made up of all the sets that are constructed by applying a finite or countable number of operations of union and intersection to the intervals of the type $A_\infty = \{x(t_i): a_i < x(t_i) \leqslant b_i,\ i = 1, 2, \ldots\}$, $C_\infty = \{x(t_i): a_i \leqslant x(t_i) < b_i,\ i = 1, 2, \ldots\}$.

type A_∞, which are related to the values of the process on this set.

Consider now the class $\mathscr{B}_T$ of events which may be interpreted as sets of realizations from $\mathscr{X}_{[0,\,T]}$ and which are generated by applying a finite or denumerable number of operations of union and intersection to the events of the type [1]

$$B_T = \{x(t),\ 0 \leqslant t \leqslant T:\ x(t_i) \in I_i,$$
$$0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T,\ \ I_i \in R^1,\ i = \overline{1, n}\}$$

It is proved in [139], that the family of finite-dimensional distributions does uniquely define the probabilities of any events belonging to $\mathscr{B}_T$. The class $\mathscr{B}_T$ is referred to as the class of Borel events on $\mathscr{X}_{[0,\,T]}$.

Since event A_T, defined in Example 4, does not belong to $\mathscr{B}_T$ for any set $\mathscr{T}$ which is dense everywhere in $[0, T]$[2], the family of finite-dimensional distributions is not enough to uniquely calculate probabilities of such events. For instance, we can generate such two random processes of one and the same family of distributions, for which the probabilities of these non-Borelian events will not be equal. Simple examples of such processes can be found in [347]. As these events are not infrequently of interest in the theory and applications, a way out of this stalemate should be found.

From the practical viewpoint, the realizations of the random process observed may deem to be continuous functions of time on the interval $[0, T]$ or have, at most, jump discontinuities there. Every function analytical in the above sense will be completely defined on $[0, T]$ if it is assigned on any countable set of points t dense in $[0, T]$ (e. g., on the set of all rational points of $[0, T]$). In this case the events of the type A_T occur if and only if the respective events A_∞ occur on the same trials. As $A_\infty \in \mathscr{B}_\infty$ then we may unumbiguously compute $\mathrm{P}(A_\infty)$ through the family of finite-dimensional distributions to finely set $\mathrm{P}(A_T) = \mathrm{P}(A_\infty)$. Assuming all realizations to be continuous or have only a finite number of ordinary discontinuities is practically not burdensome for the model within the framework of communication theory. (In stochastic process theory, there are other examples of regularity conditions to uniquely define A_T type events probability. The most general of them is the condition of existence of a separable process which is stochastically equivalent to the original one.)

The Kolmogorov theorem. In the above argument we have demonstrated that the family of finite-dimensional distributions

$$\{F(x_1^n, t_1^n),\ 0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T,\ n = 1, 2, \ldots\}$$

[1] Sets of realizations of the type B_T are named cylindrical in stochastic process theory.

[2] The sets of type A_T cannot be obtained by a countable number of operations on denumerable collection of events B_T.

may be viewed as a general stochastic model to completely describe practically all real signals and interfering effects in communication channels. Selecting various distribution families stands not for choosing an appropriate model. In the circumstances, however, the question arises of what conditions should be met for the would be chosen system of distributions to serve as the family of finite-dimensional distributions of a process. The Kolmogorov theorem gives the following answer to this [139]. The set of distributions $F(x_1^n, t_1^n)$ which satisfies the conditions [1]

$$F(x_{i1}, x_{i2}, \ldots, x_{in}; t_{i1}, t_{i2}, \ldots, t_{in}) = F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$$

$$F(x_1, x_2, \ldots, x_k; t_1, t_2, \ldots, t_k) = F(x_1, x_2, \ldots, x_k, +\infty, \ldots, +\infty; t_1, t_2, \ldots, t_k, t_{k+1}, \ldots, t_n)$$

may be adopted as the family of finite-dimensional distribution functions of a process. Here, $k < n \begin{pmatrix} 1, 2, \ldots, n \\ i1, i2, \ldots, in \end{pmatrix}$ is an arbitrary substitution.

It is obvious that for the family to be used as a stochastic model of a real process, it should be given constructively, that is by a finite number of functions with a finite number of operations on them.

1.2. Stochastic Models Based on Finite-Dimensional Distributions

Introductory remarks. As has been shown above, the stochastic model of any real fluctuation process from the communications field can be generally found as a family of finite-dimensional distribution functions (d.f.). It is this family that commonly contains complete information on the stochastic properties of the process observed. For the model to be practically sound, its finite-dimensional d.f. should be expressed by a finite number of functions of a finite array of arguments by means of some definite operations [246]. In what follows, we shall consider the cases in which the required d.f. can be arrived at with a relative simplicity the reason why the resulted models of signals and interfering effects are popular in solving applicational problems. The objective of this consideration is to display the arsenal of stochastic modeling and to refer to the literature where a more thorough treatment of the models can be found.

The normal random process. We say that $X(t)$, $0 \leqslant t \leqslant T$, is a Gaussian or normal random process if for every finite set of times

[1] These are the necessary and sufficient conditions commonly called the conditions of symmetry and consistency.

$0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T$ $(n \geqslant 1)$ its n-dimensional distribution function can be found as

$$F(x_1^n, t_1^n) = [(2\pi)^n \det \mathbf{K}]^{-1/2} \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} \exp\{(-1/2)$$
$$(y_1^n - m_1^n)' \mathbf{K}^{-1} (y_1^n - m_1^n)\} \, dy_1^n \tag{1.5}$$

where $(m_1^n)' = (m_1, m_2, \ldots, m_n)$, $m_i = m(t_i) = E[X(t_i)]$ is the mean (expectation) of the process $X(t)$ at the point $t = t_i$, $i = \overline{1, n}$; $\mathbf{K}$ is the autocovariance matrix whose elements $K_{ij} = K(t_i, \tau_j)$ are values of the autocovariance function $K(t, \tau) = E[\{x(t) - m(t)\}\{X(\tau) - m(\tau)\}]$ and $\mathbf{K}^{-1}$ is the inverse of $\mathbf{K}$. If $\mathbf{K}$ is singular, then in Eq. (1.5) $\mathbf{K} + \varepsilon\mathbf{I}$ should be substituted for $\mathbf{K}$, where $\mathbf{I}$ is the identity matrix, and pass to the limit $\varepsilon \to 0$ afterward.

We specify no conditions for $m(t)$ in Eq. (1.5), however the autocovariance $K(t, \tau)$ must be positive definite, that is for any nonzero sequence of numbers v_i, $i = \overline{1, n}$, the quadratic form

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(t_i, \tau_j) v_i v_j > 0.$$

Accordingly, the hierarchy of finite-dimensional distributions of the normal process is completely described by only two functions:

$$m(t) \text{ and } K(t, \tau), \; 0 \leqslant t, \; \tau \leqslant T \tag{1.6}$$

so that the set of these functions may be called the stochastic model of the normal random process. By selecting various functions (1.6) we in fact make a choice from a variety of models. In particular, with $m(t) = m$ and $K(t, \tau) = K(t - \tau)$, the random normal process of finite-dimensional distributions to Eq. (1.5) is stationary, that is invariant to translation in time. In this case, the check of $K(\tau)$ being positive definite is needed to approve the validity of using this function in Eq. (1.5). In the circumstances, this is an easy task: in order for the autocovariance $K(\tau)$ to be positive definite it is sufficient to require that its Fourier transform $F(f)$ (the power, or intensity, spectrum) be nonnegative and not identically zero. (This is equivalent to the assertion that $K(\tau)$ is an even function and that $|K(\tau)| \leqslant K(0)$).

The normal random process is employed for modeling mainly additive fluctuation noise. Many real signals lacking suitably simple stochastic models are also represented by the normal process. Among these signals are often found music, voice, composite signals of multichannel telephone and data transmission links, etc. The usage of the normal random process to model a wide variety of physical phenomena is valid whenever the conditions of the central limit theorem are satisfied.

The normal white noise. The stationary random process of zero mean is called a white noise process if its power spectrum is

$$F(f) \equiv N_0/2, \quad -\infty < f < +\infty$$

The autocovariance function for such a process is given by the Wiener-Khinchin theorem [162] as

$$K(\tau) = \int_{-\infty}^{\infty} F(f) \exp(i2\pi f\tau)\, df$$

$$= (N_0/2) \int_{-\infty}^{\infty} \exp(i2\pi f\tau)\, df = (N_0/2)\, \delta(\tau)$$

where $\delta(\tau)$ is the delta function. By definition, the correlation coefficient of the process is other than zero only at $\tau = 0$. The variance of the process is

$$\sigma^2 = \int_{-\infty}^{\infty} F(f)\, df = +\infty$$

In reality, the process of infinite variance does not exist. Nevertheless, the white noise model is often used to represent real broadband fluctuation noise in numerous applications.

If the output power of a filter is sought, then the following known expression is used [162]

$$\sigma_{out}^2 = \int_{-\infty}^{\infty} |K(if)|^2 F(f)\, df \tag{1.7}$$

Here $F(f)$ is the noise power spectrum at the input of the linear band-pass filter having the transfer function $K(if)$.

If $F(f)$ is a composite function of frequency, then even with a simple $|K(if)|^2$ the integration of Eq. (1.7) can offer computational difficulties. At the same time, if both $|K(if)|^2 \equiv 0$ and $F(f) = N_0/2$ for $f \in [f_l, f_u]$, then

$$\sigma_{out}^2 \approx (N_0/2) \int_{-\infty}^{\infty} |K(if)^2\, df$$

Similar argument may be presented to prove the adequacy of white noise to real broadband normal noise in problems of optimum linear filtering once the squared modulus of the transfer function is substituted by the power spectrum of a desired random signal, $F_s(f)$.

As often as not, the real broadband noise is assumed to be normal. Therefore the normality property would be desired also for the white noise introduced into the problem. However, the infinite variance of the latter makes the appropriate normal probability density void of sense, because for this process a finite dimensional distribution simply does not exist. Thus we may use a distribution of this process only in generalized sense. Indeed, white noise is referred to generalized random processes [75] whose description may be inferred from the following argument.

Let $X(t)$ be a random process. Any linear device enables one to observe or measure not the values of the process $X(t)$ at a time t_0, but the values of the linear functional

$$\Phi(h, t_0) = \int_T h(t_0, \tau) X(\tau) d\tau$$

where $h(t_0, \tau)$ is a function which characterizes the given device. If only the readings of the device are of interest, then we may entirely avoid discussing the process $X(t)$. Observe, the finite-dimensional distribution of the process may not exist.

Reference [75], for example, suggests the following way of describing a generalized random process. Let $\mathcal{H}$ be a set of functions h each of which possesses all its derivatives and is zero outside some finite interval. Let $\Phi(h)$ be a linear functional on $\mathcal{H}$. Then a generalized random process is deemed to be specified if for any h of $\mathcal{H}$ the random variable $\Phi(h)$ is given. A generalized process $X(t)$ is called normal if for any h of $\mathcal{H}$ the respective random variable $\Phi(h)$ is normally distributed. A generalized process $X(t)$ is called stationary if for any τ the probabilities that the inequalities $a < \Phi(h) < b$ and $a < \Phi[h(t + \tau)] < b$ are satisfied are equal, that is the process $\Phi(\tau) = \Phi[h(t + \tau)]$, $-\infty < \tau < +\infty$, is invariant to translation in time.

The term 'normal white noise' has a broad usage. However, it should be understood in the above generalized sense: white noise is a normal process if being applied to the input of a deterministic linear system it produces at the output also a normal process.

The narrowband normal process. A stationary random process is said to be frequency bounded about the central frequency f_0 if its power spectrum for positive frequencies is zero outside the interval $(f_0 - \Delta, f_0 + \Delta)$ and Δ is subject to inequality $\Delta < f_0$. If $\Delta \ll f_0$, the process is termed narrowband. The narrowband stationary normal process $X(t)$ may be represented [162] as

$$X(t) = A(t) \cos 2\pi f_0 t + B(t) \sin 2\pi f_0 t$$

where $A(t)$ and $B(t)$ are the stationary, statistically related normal processes whose power spectra $F(f_0 + f) + F(-f_0 + f)$ are essentially zero outside the interval $(-\Delta, \Delta)$. If the expectation $E[X(t)] \equiv 0$, then any finite dimensional joint distributions of the processes $A(t)$ and $B(t)$ are completely defined by their autocovariances $K_A(\tau)$ and $K_B(\tau)$ and cross-covariance $K_{AB}(\tau)$. A more detailed treatment of the normal narrowband process may be found in [162], where the definition is given via the Hilbert transform and in [315], where to the purpose the representation in the frequency domain is involved.

To sum up, we should point out that narrowband normal processes are the most frequently used models of signals and disturbances in radio communication and multichannel telecommunication.

The Markov process. An important property of the Markov process is in that its hierarchy of finite dimensional distributions can be described by only a small number of characteristics. Then the probabilities of various events related to the process are an easy matter of mathematical routine. The notion of the process has been introduced by Markov, a Russian mathematician, as a generalization of sequential events making up a chain. The general features of certain Markov process classes were brought out by Kolmogorov in 1931. The mathematical theory of Markov processes embeds some of his ideas presented in [416].

The Markov process is a stochastic process in which the probabilities of events related to the values of the process in the future depend only on the value $X(t)$ at the present time t and are independent of the values $X(\tau)$ in the past, that is at $\tau < t$. Thus for any times $t_0 < t < s$ and a given value of $X(t)$, the random variables $X(t_0)$ and $X(s)$ are statistically independent. It may be said that the knowledge of the "present" makes the "past" and the "future" stochastically independent of each other [430].

That Markov models can adequately describe thermal noise in electric circuits is shown, for instance, in [315]. In the reference, a technique is given to construct a vector Markov process which includes as a component the original non-Markovian process being modeled, and certain examples are displayed of Markov models being employed in numerous problems of communication. Markov processes as stochastic models of noise and signals are also discussed in [253, 262, 466].

It is a relatively easy matter to describe for a Markov process its family of finite-dimensional distributions. By definition of the Markov process, for any times t' and t such that $t' < t$

$$\begin{aligned} \mathrm{P}\{X(t) \leqslant x \mid X(\tau) = x'(\tau),\ \tau \leqslant t'\} \\ &= \mathrm{P}\{X(t) \leqslant x \mid X(t') = x'(t')\} \\ &= F(x;\ t \mid x';\ t') \end{aligned} \tag{1.8}$$

where $F(x;\ t \mid x';\ t')$ is the transition probability distribution function which defines the probability for the process initially having the value x' at the time t' to transfer to a value less than x at the moment t. For any set of times $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T$, the distribution function of any random process may be given in the form

$$\begin{aligned} F(x_1^n, t_1^n) = F(x_1, t_1)\, F(x_2, t_2 | x_1, t_1)\, F(x_3, t_3 | x_1^2, t_1^2) \ldots \\ \times F(x_n, t_n | x_1^{n-1}, t_1^{n-1}) \end{aligned}$$

For the Markov process, due to Eq. (1.8), the above equation may be rewritten (so-called factorization of the multivariate d.f.) as

$$F(x_1^n, t_1^n) = F(x_1, t_1) \prod_{k=2}^{n} F(x_k, t_k | x_{k-1}, t_{k-1}) \tag{1.9}$$

From Eq. (1.9) it follows that any finite dimensional distribution of the Markov process is completely defined by two functions:

$$F(x, t) \text{ and } F(x, t \mid x', t'), \ -\infty < x, \ x' < +\infty, \quad 0 \leqslant t' < t \leqslant T \tag{1.10}$$

Thus, the functions (1.10) completely describe the model of the Markov process. Accordingly, selection of various pairs of these functions may be interpreted as choosing various stochastic models of the Markov class.

It should be noted that $F(x, t)$ is a one-dimensional d.f. of general form having no additional constraints. The transition probability d.f. $F(x, t \mid x', t')$ is not to choose in an arbitrary fashion. In particular, only that conditional probability distribution function may be chosen for a transition probability distribution for which the equation of Chapman and Kolmogorov (or, equivalently, the Smoluchowski equation) holds. In terms of finite-dimensional d.f., this equaion may have the form [162]:

$$w(x, t|x_0, t_0) = \int_{-\infty}^{\infty} w(x, t|x', t')\, w(x', t'|x_0, t_0)\, dx', \quad t_0 < t' < t \tag{1.11}$$

Consequently, selection of a Markovian model suggests that Eq. (1.11) should be solved. For more narrow classes of Markov processes this equation may be reduced to linear differential equations (the Markov processes of diffusion) or to integro-differential equations (step-type processes and diffusion processes having jumps). General methods to solve these equations are as yet to find. An exclusion presents a particular case of normal Markov processes [246]. Thus, the seemingly natural way of constructive description of stochastic models by a finite number of functions of finite numbers of time and space arguments is not the best one and not always objective-bound. Indeed, already for the functions of two time and two space arguments (the function $F(x, t \mid x', t')$ for Markovian processes), we would have to bump into purely mathematical difficulties which are still to overcome. Another sufficiently different and often very convenient approach to description of Markov processes is based on stochastic differential equations and will be discussed in Sec. 1.3.

Processes with independent increments. The Markov processes with independent increments form a class of processes which are extremely important from both the theoretical and practical points of view. One of them, the Poisson process, serves as a very accurate model for many physical situations, such as radioactive desintegration, emission of electrons, occurrence of telephone calls, and various other events. Another, the Wiener process, is very important in the

theoretical development of stochastic system modeling as with Kalman filtering and is also used for a rigorous treatment of stochastic differential and integral equations.

A stochastic process $X(t)$ with independent increments is a process whose increments over two non-overlapping intervals are independent random variables, that is if for any n and any set of times $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant T$ the random variables $X(t_1)$, $X(t_2)-X(t_1)$, ..., $X(t_n)-X(t_{n-1})$ are independent. The family of finite-dimensional distribution functions of this process will be completely defined if we specify only two functions

$$F(x, 0) \text{ and } F(z;\ t_1,\ t_2),\ -\infty < x,\ z < +\infty,$$
$$0 \leqslant t_1 < t_2 \leqslant T \tag{1.12}$$

where $F(z; t_1, t_2)$ is the d.f. of random variable $Z(t_1, t_2) = X(t_2) - X(t_1)$. To prove, for any finite-dimensional distribution of the process the following representation is valid [246]

$$F(x_1^n, t_1^n)$$
$$= \int_{-\infty}^{\infty} w(z_0, 0) \prod_{k=1}^{n} \varepsilon\left(x_k - \sum_{i=0}^{k} z_i\right) \prod_{k=1}^{n} w(z_k, t_{k-1}, t_k)\, dz_0^k$$

Here $\varepsilon(z) = 1$ for $z > 0$ and $\varepsilon(z) = 0$ for $z < 0$ and $t_0 = 0$. The distribution function $F(z; t_1, t_2)$ should be chosen such that the following equation must hold

$$F(z; t_1, t_3) = \int_{-\infty}^{\infty} F(z-u, t_1, t_2)\, dF(u; t_2, t_3)$$

The characteristic function of the distribution $F(z; t_1, t_2)$, which facilitates the choosing of this distribution, is described in [246].

Hence, the stochastic model of the process with independent increments may be understood as the functions (1.12). If $F(z; t_1, t_2)$ depends on the difference of the times considered only, then the process is termed uniform.

The Poisson process. Let $X(t)$ be a process with independent increments, then $X(t)$ is called a Poisson process if the following assumptions are made: (1) $X(0) = 0$, $X(t)$ is a nondecreasing step function which increases with a jump 1 at each discontinuity and is constant (the realizations are constant) between any neighbouring jumps; (2) the probability that exactly n jumps (points) occur on the interval (t_1, t_2) is given by the Poisson distribution:

$$\mathrm{P}\{X(t_2) - X(t_1) = n\} = \exp[-\lambda(t_2 - t_1)]\, \lambda^n/n!$$

where λ is a positive constant. The probability that exactly n points occur in the time interval $(0, t)$ depends on both n and t, but not on the position of the interval $(0, t)$. For example, if we set $t = t_2 - t_1$

in the above equation, then

$$P\{X(t)=n\}=\exp(-\lambda t)\lambda^n/n!$$

The Poisson process is a simplest tool to model overloaded traffic in multichannel communication lines, impulse arrivals of random impulse noise in radio channel, message array inputs in stochastic multi-access systems, etc.

The Wiener or Brownian motion process is another important process with independent increments. Let $X(t)$ be a process with independent increments. Then $X(t)$ is a Wiener process if the following conditions are satisfied: (1) $X(t)$ is a normal process for every t; (2) $E[X(t)] = 0$ and $K(t_1, t_2) = \min(t_1, t_2)$ for $0 \leqslant t_1, t_2 \leqslant T$. The Wiener process serves as a key element in the development of stochastic integral and stochastic differential equation calculus (see Sec. 1.3).

1.3. Modeling with Stochastic Differential Equations

Introductory remarks. In the following, we shall shortly dwell on another constructive approach to random process, which also completely defines the stochastic properties of the process. This approach enables us to efficiently solve the problems of both synthesis of optimal communication systems and analysis of these systems under random noise. In addition, it features a dynamical model that enables the solution of complicated integral equations defining the impulse response of optimal linear device to be avoided by reducing this stage to a problem of modeling.

This approach is based on the properties of continuous Markov processes and the concept of state variable. The practical importance of the approach has risen considerably after the works [400, 401] had been published, in which an analogous technique to define random processes was used to advantage in solving the linear filtering problem. The novelty in the statement of this problem was in that all the processes concerned (desired signals and noise) were represented by differential equations rather than by covariance functions (finite-dimensional distributions) as is the case in the Kolmogorov-Wiener theory of linear filtering. The new approach proved to be fruitful from the viewpoint of practical realization for the optimal filters. Thus, having confined ourself to the normal and Markov models of message and noise and defined them by a set of differential equations, the differential equations have been obtained to optimally estimate the message being filtered. The last equations are such that they can be easily solved on an analog or digital computer. Observe, that in the Kolmogorov-Wiener theory, the impulse response of an optimal filter is found by solving the nonhomogeneous integral equation. Practical utility of the state-variable equation method in

describing random processes has been also illustrated by another approach to the problem of filtering, which has been first suggested by Stratonovich [252]. Systematic expansion of this method as applied to the problems of analog communication is developed in references [466, 484, 494].

The Markov process. The state-variable approach assumes the treatment of the process considered $X(t)$ as a signal at the output of some, generally nonlinear, dynamic system described by the following stochastic differential equation[1]

$$dX(t)/dt = F[X(t),\ t] + G[X(t),\ t]\,n(t),$$
$$X(t_0) = X_0,\quad t \geqslant t_0 \tag{1.13}$$

where $n(t)$ is the normal white noise impressed across the input of the system; $F(x, t)$ and $G(x, t)$ are the functions of x and t, nonlinear with respect to x and providing memoryless transformations (in many problems of communication, $G(x; t)$ is assumed to depend on t only, thus using Eq. (1.13) in the simplified form); X_0 is the value $X(t)$ takes on at t_0, when the observation begins (X_0 may be a random variable). For Eq. (1.13) to completely describe the process $X(t)$, X_0 must be supplied with its distribution function or probability density $w(x_0)$. The conditions which the process $X(t)$ of Eq. (1.13) should satisfy to become a Markov process are presented in [466].

It has been shown (see, for instance, [316]) that the conditional probability density $w(x, t \mid x', t')$, $t' < t$, which completely describes the stationary Markov process $X(t)$ of Eq. (1.13), obeys the following partial differential equation of the diffusion type, known as the Fokker-Planck-Kolmogorov equation,

$$\frac{\partial}{\partial t} w(x, t)|x_0, t_0) = -\frac{\partial}{\partial x}[F(x, t)\,w(x, t|x_0, t_0)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\frac{N_0}{2} G^2(x, t)\,w(x, t|x_0, t_0)\right]$$

at the initial condition $w(x, t_0 \mid x_0, t_0) = \delta(x - x_0)$. Thus, when the distribution densities $w(x_0, t_0)$ and $w(x, t \mid x_0, t_0)$ are given, the total family of finite-dimensional distributions for the Markov process can be constructed.

The normal random process. A mathematical representation of a normal random process having the power spectrum of the rational fraction type may be obtained making use of a filter with the transfer

---

[1] The description of random process at the output of a dynamic system by the differential equation (1.13) we shall call the *direct* tecnhique in contrast to the indirect one which relates the stochastic models of input and output for the system. The topic will be treated in more detail in Ch. 2.

function

$$K(s) = (b_1 s^{m-1} + b_2 s^{m-2} + \ldots + b_{m-1}s + b_m)/$$
$$(s^m + a_1 s^{m-1} + \ldots + a_{m-1}s + a_m) \tag{1.14}$$

when normal white noise $n(t)$ is applied to its input. This filter is described by the following differential equation, relating the input, $\tilde{n}(t)$, and the output, $S(t)$, processes,

$$\frac{d^m S(t)}{dt^m} + a_1 \frac{d^{m-1}S(t)}{dt^{m-1}} + \ldots + a_m S(t)$$
$$= b_1 \frac{d^{m-1}\tilde{n}(t)}{dt^{m-1}} + b_2 \frac{d^{m-2}\tilde{n}(t)}{dt^{m-2}} + \ldots + b_m \tilde{n}(t), \quad t \geqslant t_0 \tag{1.15}$$

where a_i, b_i, $i = \overline{1, m}$, are the constant parameters of the filter.

An alternative representation of the process $S(t)$ may be obtained by transforming the above differential equation of mth order into m simultaneous differential equations of first order.

$$\begin{aligned}
dX_1(t)/dt &= -a_1 X_1(t) + X_2(t) + b_1 \tilde{n}(t) \\
dX_2(t)/dt &= -a_2 X_1(t) + X_3(t) + b_2 \tilde{n}(t) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
dX_{m-1}(t)/dt &= -a_{m-1} X_1(t) + X_m(t) + b_{m-1} \tilde{n}(t) \\
dX_m(t)/dt &= -a_m X_1(t) + b_m \tilde{n}(t)
\end{aligned} \tag{1.16}$$

where

$$X_1(t) = S(t) \tag{1.17}$$

To prove that this system of equations is equivalent to Eq. (1.15), each of the simultaneous equations should be differentiated an appropriate number of times and then substituted into the previous equation.

Equations (1.16) are called the equations of state variable. To write the system in the vector form, we introduce the vector of state variable $\mathbf{X}_1^m(t) = [X_1(t), X_2(t), \ldots, X_m(t)]^T$. Then we have for Eqs. (1.16)

$$d\mathbf{X}_1^m(t)/dt = \mathbf{F}\mathbf{X}_1^m(t) + \mathbf{G}\tilde{n}(t) \tag{1.18}$$

where

$$\mathbf{X}_1^m(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_m(t) \end{bmatrix}, \mathbf{F} = \begin{bmatrix} -a_1 & 1 & 0 \ldots 0 \\ -a_1 & 0 & 1 \ldots 0 \\ -a_1 & 0 & 0 \ldots 0 \\ \ldots & \ldots & \ldots \\ -a_1 & 0 & 0 \ldots 1 \\ -a_1 & 0 & 0 \ldots 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

Modulated signal and additively admixed noise. Let $S(t)$ be a signal to be transmitted through the channel where it mixes with an additive normal noise $n(t)$. With phase modulated carrier, the process observed at the output of the channel may be represented as

$$X(t) = A \sin [2\pi f_0 t + \beta S(t)] + n(t), \quad t \geqslant t_0 \tag{1.19}$$

Below we shall demonstrate how this expression can be written as the stochastic differential equation of the type of Eq. (1.13) but in the vector form.

Assume that $S(t)$ is a component of a vector process satisfying the system of Eqs. (1.16) and (1.17). This implies the normal process of rational fraction power spectrum due to the transfer function of Eq. (1.14). Next, introduce a process $\tilde{X}(t) = \int_{t_0}^{t} X(\tau)\, d\tau$ for which $d\tilde{X}(t)/dt = X(t)$. As a result, Eq. (1.19) may be written as the stochastic differential equation

$$d\tilde{X}(t)/dt = h[S(t), t] + n(t), \quad t \geqslant t_0 \tag{1.20}$$

Further on, Eq. (1.18) and Eq. (1.20) may be treated as the system of simultaneous stochastic equations

$$\begin{aligned} d\tilde{X}(t)/dt &= h[X_1(t), t] + n(t) \\ d\mathbf{X}_1^m(t)/dt &= \mathbf{F}\mathbf{X}_1^m(t) + \mathbf{G}\tilde{n}(t) \end{aligned} \tag{1.21}$$

which, obviously, can be rewritten in the vector form as

$$d\mathbf{Y}_1^{m+1}(t)/dt = \tilde{\mathbf{F}}[\mathbf{Y}_1^{m+1}(t), t] + \tilde{\mathbf{G}}\tilde{\tilde{n}}(t) \tag{1.22}$$

where $\mathbf{Y}_1^{m+1}(t) = [\tilde{X}(t), X_1(t), \ldots, X_m(t)]^T$ is the column matrix, T stands for the transpose; $\tilde{\tilde{n}}(t) = [n(t), \tilde{n}(t), \ldots, \tilde{n}(t)]^T$,

$$\tilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & b_1 & 0 & \ldots & 0 \\ 0 & 0 & b_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & b_m \end{bmatrix}, \quad \tilde{\mathbf{F}}[\mathbf{Y}_1^{m+1}(t), t] = \begin{bmatrix} h[X_1(t), t] \\ \mathbf{F}\mathbf{X}_1^m(t) \end{bmatrix}$$

and $n(t)$ and $\tilde{n}(t)$ are independent Gaussian noises.

A more detailed discussion concerning the representation of signals and noise by the state-variable approach may be found in [466, 484, 495].

1.4. Stochastic Time-Spatial Models of Communication Channels

Classification of models. The majority of real transmission channels may be visualized as media with random inhomogeneities, which as a random field may be described by a stochastic model. The theory of wave propagation in media with random inhomogeneities is well developed, mainly in works of Soviet scientists [5, 15, 98, 231, 258, 287].

In studying wave channels with random inhomogeneities (dispersive channels), recourse is widely had to phenomenological approach treating the propagation of waves as multipath (multibeam) channel phenomena. This approach is based on the experimental evidence advocating that the major role in forming the scattered field at the receiver is played by the so-called "glaring" spots covering only a relatively small part of the scattering surface, or volume, of the scattering object. Contributions in the scattered field at the receiver may be made by numerous spatially diversed scattering regions. If the scattering surface somewhat changes its orientation or/and shape with time, the glaring spots also move and change in intensity.

When every beam reverberates more than once prior to coming to a receiver, the phenomenon is called multiple scattering. In some cases the effect of multiple scattering may be neglected [215, 442]. Single scattering gives rise to the model of parallel propagation, and multiple scattering to the sequential-parallel propagation model. With single scattering, the beams at the receiving end may be treated as independent, and the respective antenna field may be described as

$$s(t, \mathbf{r}) = \sum_{k=1}^{N} s_k(t, \mathbf{r}) \tag{1.23}$$

where $s_k(t, \mathbf{r})$ is the process describing the kth random path.

In channels of practical interest the number of glaring spots is large. As the contribution of every spot in the total field is usually small and the contributions may be deemed independent, the central limit theorem may be applied so that the field at the channel output may be treated as normal.

The stochastic description of the channel with the normality assumption proves most simple and reduces to determination of the first two moments of the field—the regular component $S(t, \mathbf{r})$ and the autocorrelation function $K_s(t, t+\tau; \mathbf{r}, \mathbf{r}+\Delta\mathbf{r})$. There exists a variety of works [97, 129, 215, 297, 322] devoted to determination of regular components and correlation functions at the output of physically different channels.

The investigation is much more complicated in the case of multiple scattering [127, 135, 258, 442]. The field, as for single scattering, may be represented as a sum with the difference that the terms

may not be assumed independent. Thus the rigorous solution of the wave differential equation, subject to boundary effects and random properties of the medium, is extremely difficult. A way out towards the stochastic characteristics of the receiving field is the model of the channel involving a final set of linear time-spatial filters having the characteristics $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_N$. Every such filter takes part in shaping the field at the reception region both directly and via all or part of the rest filters [135]. As the linear operators $\mathbf{L}_k$, we shall use the complex transfer functions $K_k(t, f, \mathbf{r})$. Then, in parallel propagations,

$$K(t, f, \mathbf{r}) = \sum_{k=1}^{N} K_k(t, f, \mathbf{r}) \tag{1.24}$$

and in sequential propagation,

$$K(t, f, \mathbf{r}) = \prod_{k=1}^{N} K_k(t, f, \mathbf{r}) \tag{1.25}$$

A separate discussion will be given to the case of sequential-parallel propagation, for which

$$K(t, f, \mathbf{r}) = \sum_{k=1}^{N} \prod_{l=1}^{L_k} K_{lk}(t, f, \mathbf{r}) \tag{1.26}$$

Stochastic models of fading channels. In constructing the model of fading channels, the number of scattering objects, N, which form a total signal at the receiver, is usually assumed to be large [127, 128, 131, 132, 199, 267, 442, 444]. However, in sequential-parallel propagation this condition would be insufficient to find the distribution of the random variable of Eq. (1.26) in the limit as $N \to \infty$.

Let us consider first two basic situations: the additive case of Eq. (1.24) and the multiplicative of Eq. (1.25).

For the additive situation, the conditions of the central limit theorem may be deemed valid as $N \to \infty$. This makes it possible to treat the transfer function as a complex random variable, namely,

$$K(t, f, \mathbf{r}) = X(t, f, \mathbf{r}) + iY(t, f, \mathbf{r})$$

Its quadrature components $X(t, f, \mathbf{r})$ and $Y(t, f, \mathbf{r})$ generally depend on each other and have arbitrary means m_X and m_Y and variances $\sigma_X^2 \neq \sigma_Y^2$.

Kotel'nikov used the model of stationary normal process to describe additive noise in the channel [151]. Siforov [245] used the same model to describe fadings in radio channel (the Rayleigh distribution of amplitudes and the uniform distribution of phases). Experimental data evidenced the validity of the generalized Rayleigh model for channels in various frequency bands [5]. This model has also been used in works [130, 492, 493].

Korzhik and Fink [147] discussed a more general model of linear nonstationary stochastic channel. They questioned whether it was correct to approximate continuous channels by the developed model

and considered the linear stochastic channel as approximated by linear transforms with a difference kernel and by a multiplicative transform.

The experimental data, while suggesting that the normal approximation of stochastic fields is valid in many frequency bands, is in poor agreement with the generalized Rayleigh model. In particular, some fading of signal has been observed to be more pronounced than the Rayleigh's one [5, 182, 444, 445] and the phase distribution of the stochastic field component has been found to be nonuniform [212, 319, 320].

Hogt was the first to study the model of normal random vector with independent quadrature components of different variances, the phase being distributed nonuniformly [388]. Nakagami used this model to describe the distribution of signal amplitude. In work [319] this model in the generalized form (three-parameter distribution [132]) was employed to describe the distributions of amplitude and phase dispersed by rough surfaces.

A more complicated model, later called the four-parameter model, was analyzed in [132, 320, 442, 444]. This model was utilized to describe the stochastic laser channel [182] and to investigate radio communication channels with statistically rough surfaces [215]. In [212] the model is introduced from the theoretical analysis of wave propagation in a stochastically inhomogeneous medium, and a technique is developed to obtain the parameters of the model on experimental data.

The four-parameter distribution of modulus K has the form:

$$w(\gamma) = \sum_{n=0}^{\infty} \frac{R^n}{n!} \sigma^{2n} \frac{\partial^{2n}}{\partial m_I^n \partial m_{II}^n} \left[\frac{\gamma}{\sigma^2} \exp\left(-\frac{\gamma^2 + m_I^2 + m_{II}^2}{2\sigma^2} \right) \times I_0 \left(\frac{\gamma (m_I^2 + m_{II}^2)^{1/2}}{\sigma^2} \right) \right] \tag{1.27}$$

where

$$m_I = (m_X + m_Y)/\sqrt{2}, \quad m_{II} = (m_X - m_Y)/\sqrt{2} \tag{1.28}$$

$$\sigma^2 = (\sigma_X^2 + \sigma_Y^2)/2, \qquad R = (\sigma_Y^2 - \sigma_X^2)/(\sigma_Y^2 + \sigma_X^2) \tag{1.29}$$

There also exist other forms for this distribution [132, 135, 442, 444]. A number of valuable formulas related to the four-parameter distribution of modulus and argument of the complex transfer function may be found in [132, 319, 442, 444].

The experimental data indicates the four-parameter distribution and its various specific cases cover a broad class of communication channels [5, 127, 128, 445, 492]. The solution of stochastic wave equation for different mechanisms of propagation leads also to this model [97, 98, 215, 287, 319, 442]. It should be emphasized that the

four-parameter distribution of modulus is always due to nonstationarity of the channel, which is typical of most channels in actual use.

The two-parameter m-distribution for the modulus of sum (1.24), suggested in [444] gives a satisfactory approximation to the four-parameter distribution of amplitudes. The model of m-fading channel was also used in [74, 87]. The work [444] yet to have been published, Konoploeva had suggested this distribution for the integer-valued parameter $2m$, working on the experimental data from several radio relay links [149].

In [194] it was suggested to model fadings (multiplicative noise) in radio communication channels as a product of two random processes, one describing slow fluctuations of the signal and the other fast. These fadings were called bimultiplicative [276]. In this reference, a specific model of these fadings was suggested, which have a distribution of the product of two independent Rayleigh random variables (see [162]). The distribution describes the fadings that are deeper than the Rayleigh fadings. In the majority of works, the lognormal distribution is invoked to describe the slow fluctuations of the envelope in radio channels. Some works (e.g., [294]) use the gamma-distribution to represent fluctuations of signal and noise intensities. In [154] the stochastic channel is modeled as a link applying amplitude and phase modulations to the transmitted signal, and then various specific cases of the model are discussed.

If we write the complex transfer function of the kth partial filter in Eq. (1.25) as $K_k = \exp \alpha_k \exp i\varphi_k$, then, summed up, $K = \exp(\alpha + i\varphi) = \gamma \exp i\varphi$. For the quantities $a = \sum_{k=1}^{n} a_k$ and $\varphi = \sum_{k=1}^{n} \varphi_k$, the conditions of the central limit theorem hold as $N \to \infty$, enabling us to treat them as gaussian random variables. Hence, the modulus $\gamma = \exp a$ has a lognormal distribution. When fluctuations of signal intensity are deep, the distribution of the phase may be approximated by the uniform distribution over the interval $[-\pi, +\pi]$ [258].

In principle, the situations are possible that lead to sufficiently different distributions for the variable K, being defined according to Eq. (1.26). Computer simulations display that the form of distribution for $|K|$ at $N < 10$ and the condition for the parameters of the composite filters being random but time-invariant are decided mainly by multiplicative rather than additive nature of the relationship between the terms. When the parameters of the filters vary with time at random, the multiplicative effect becomes weaker. As N increases, the additive effect of terms also increases. With a limited number of terms, the argument of K may be deemed to have a uniform distribution [162].

Models based on the correlation properties. Stochastic models of communication channels based on the correlation properties were

considered in [128, 322]. In description, a time-spatial channel features the relation between the field correlation intervals with respect to frequency F_k, time τ_k and space ρ_k, as well as parameters T_s, F_s, and R (the spatial spread of the field being analyzed at the reception locale).

The duration T and the spectrum bandwidth F of a signal at the output of a channel, that is the parameters of an analyzed signal at the receiving end, we shall define as $T = T_s + \xi_{\max}$, $F = F_s + f_{fad,\max}$ where $\xi_{\max} \approx 1/F_k$ is the interval of the signal time spread (memory of the channel) due to the deviation of response characteristic from the ideal one or due to the difference of impulse response from the delta function (because of multipath wave propagation, nonlinearity of phase-frequency characteristic, etc.); $\Delta f_{fad,\max} \approx 1/\tau_k$ is the interval of the frequency signal dispersion (or the bandwidth of the power spectrum of the time-variable fadings) stemming from the change of channel parameters with time or from mutual translation of the regions where the signal is shaped and then received; T_s and F_s are the parameters of the signal at the channel input [150].

Single-path channel with additive noise. If the conditions $T_sF_k \gg 1$, $F_s\tau_k \gg 1$ are satisfied, then the model should withhold only one propagation path, which would correspond to the *single-path channel model*.

If we pretend the channel parameters being time-invariant over the analysis interval, that is $h(t, \xi, \mathbf{r}) = h(\xi, \mathbf{r})$, then assuming $h(\xi, \mathbf{r}) = \gamma(\mathbf{r})\, \delta[\xi - \tau(\mathbf{r})]$ and utilizing the filtering property of δ-function, we shall have at the channel output

$$Z(t, \mathbf{r}) = \gamma(\mathbf{r})\, u[t - \tau(\mathbf{r})] + \xi(t, \mathbf{r}) \tag{1.30}$$

If the signal at the input, $u(t)$, and parameters of the channel, γ (scale) and τ (delay time), are known in advance and the additive noise $\xi(t, \mathbf{r})$ is a stationary normal random process whose autocovariance is given, then we have the model of nondisturbing ("ideal") channel with the additive normal noise.

This model modifies somewhat when gain γ and delay τ are assumed to be known functions of time. It will supply a fair description of many wire-communication channels, radio channels for the line-of-sight communication, and slow fading radio channels, that is of all those situations when the values of γ and τ can be predicted exactly.

For the narrowband signal, the following representation is more convenient

$$Z(t, \mathbf{r}) = \gamma(\mathbf{r})\, \{\cos\theta(\mathbf{r})\, u[t - \tau_s(\mathbf{r})] \\ - \sin\theta(\mathbf{r})\, \hat{u}[t - \tau_s(\mathbf{r})]\} + \xi(t, \mathbf{r}) \tag{1.31}$$

where $\hat{u}(t)$ is the Hilbert transform of $u(t)$, τ_s is the average delay of signal, θ is the phase shift of signal in the channel with respect to high frequency.

If γ and τ are known and θ is a random phase uniformly distributed over the interval $(-\pi, \pi)$, then we will have the model of a channel with a random phase and additive normal noise. If in Eq. (1.31) both the phase θ and gain γ are of random nature, then we have the model of a single-path fading channel, for which

$$Z(t, \mathbf{r}) = x(t, \mathbf{r})\, u(t - \tau_s) - y(t, \mathbf{r})\, \hat{u}(t - \tau_s) + \xi(t, \mathbf{r}) \tag{1.32}$$

where $x(t, \mathbf{r}) = \gamma(t, \mathbf{r}) \cos\theta(t, \mathbf{r})$

$$y(t, \mathbf{r}) = \gamma(t, \mathbf{r}) \sin\theta(t, \mathbf{r})$$

The quantity τ_s is usually assumed to be known. The model of a stochastic channel in which the moment of arrival is not known is treated in [492] and [493].

Multipath channel with additive noise. Assuming that the channel parameters do not change with time and the impulse response of the channel is given by

$$h(\xi, \mathbf{r}) = \sum_{k=1}^{N} \gamma_k(\mathbf{r})\, \delta[\xi - \tau_k(\mathbf{r})]$$

where γ_k, τ_k are the gain and delay in the kth path, respectively, we get for the signal at the output of the channel

$$Z(t, \mathbf{r}) = \sum_{k=1}^{N} \gamma_k(\mathbf{r})\, u[t - \tau_k(\mathbf{r})] + \xi(t, \mathbf{r}) \tag{1.33}$$

If the input signal $u(t)$ and the channel parameters γ_k, τ_k are known in advance, and the additive noise is Gaussian and stationary, then we will have the model of multipath channel having fixed parameters.

For narrowband signals, more convenient is the following model:

$$Z(t, \mathbf{r}) = \sum_{k=1}^{N} \gamma_k(\mathbf{r}) \{\cos\theta_k(\mathbf{r})\, u[t - \tau_{s,k}(\mathbf{r})] - \sin\theta_k(\mathbf{r})\, \hat{u}[t - \tau_{s,k}(\mathbf{r})]\} + \xi(t, \mathbf{r}) \tag{1.34}$$

where θ_k is the phase shift in the kth path, $\tau_{s,k}$ is the average delay of a signal in the kth path. If we pretend that the phase θ_k is a random variable, then the model of multipath channel with normal additive noise and random phase results, enabling us to describe many actual channels of radio communication.

For the model of a multipath fading channel (satisfactorily approaching to the majority of radio communication channels), the

signal at the output has the form

$$Z(t, \mathbf{r}) = \sum_{k=1}^{N} x_k(t, \mathbf{r})\, u(t - \tau_{s,k}) - y_k(t, \mathbf{r})\, \hat{u}(t - \tau_{s,k}) + \xi(t, \mathbf{r}] \tag{1.35}$$

where
$$x_k(t, \mathbf{r}) = \gamma_k(t, \mathbf{r}) \cos \theta_k(t, \mathbf{r})$$
$$y_k(t, \mathbf{r}) = \gamma_k(t, \mathbf{r}) \sin \theta_k(t, \mathbf{r})$$

The model given by Eq. (1.34) has been used to optimally detect binary wideband signals in a short-wave channel from the white noise background (the Rake system [455]). The model of Eq. (1.35) has been used to develop the sequential modem in the system SIIP [132, 134, 135].

If in Eqs. (1.33) through (1.35) the number of paths N tends to infinity and the values of delay form a continuum, the channels in the resulted models will have the property of continuum multipath. In the general case, these channels are stochastic and involve channel memory and dispersion, wave scattering and additive noise.

1.5. Measuring the Statistical Parameters of Signals and Channels in Communication Systems

Introductory remarks. The mathematical models discussed above often give but an idealized representation of how actual signal and channels behave in communication systems. In the circumstances the significance of the experimental verification could hardly be overemphasized. Actually, whether a given model fits the actual signal or channel well or not can be asserted by measuring the respective statistical parameters. The results of the measurement give an indication of quality with which the communication system performs. Therefore, several recent regulations display the requirements to communication systems and characteristics of the signals and channels used.

The last decade has seen the intensive development of measurement theory and engineering involving a variety of devices, algorithms, and techniques. The theory and techniques used to measure various characteristics of random processes are discussed in a series of books [125, 190, 191, 421].

Devices to measure the correlation functions, power spectrum, mean value, variance, average power, probability distribution and density functions of random processes become indispensable tools in research and development work. They are employed for production and process monitoring as well. In addition to these general characteristics, measurement of suitable statistical parameters

become more and more popular with multichannel communications involving randomly changed parameters of signals and channels.

The goals of measuring the statistical parameters in communications are many and diverse: to evaluate the average power of signals, unweighted and psophometric noise; to analyze the density functions and spectra of signals and noise; to compare the attenuation ripple of the channel with the rated values; to estimate the interchannel interference in multichannel systems; to estimate the intensity of pulse noise and short duration level interruptions in data transmission systems; to obtain the parameters of overshoots above the threshold for signals and noise; to analyze noise from switching equipment; to estimate the fidelity of data transmission in a digital form; to investigate the correlation of equipment-section and channel parameters; to determine the characteristics of signal phase fluctuations; to verify the identity of two channels; to measure the crosstalk attenuation in multichannel communication systems without interrupting their operation; to locate the failed section in the cable line, etc.

Measuring the statistical parameters. Random process measuring is based on the procedures used to measure any signal—conversion of the signal, reproduction of a unit, comparison of the signal with the unit, and recording the result of comparison. Statistical parameters differ in the measurement procedure in the following:

It is sound (and at times mandatory) to hold the realizations of the random process measured to provide for their manyfold reproducing.

An averaging operation is a must dictated by the very essence of a statistical parameter.

The true results may not be obtained unless the amount of statistical data is sufficiently large.

To estimate the measurement errors a higher order statistics is needed than that being measured.

The instruments employed to measure statistical parameters require an elaborate technique for their testing and verifying.

Measurement of mean power. Mean power of a signal or noise, and signal-to-noise ratio are the commonly measured parameters in wire and radio communication, and in radio and TV broadcasting. For RF circuits and channels of communication systems, the rated values are issued for the mean power of unweighted noise, and the mean power of noise psophometrically weighted; the levels of signal power and signal-to-noise ratio are measured. In PCM systems, the mean power of quantization and signal-to-noise ratio are measured. In radio broadcast, the information about the average power of antenna signal is of value. An estimate of radio communication system noise immunity is arrived at by measuring the noise power and signal power at various types of modulation.

The average power of an ergodic random process $X(t)$ dissipated in a resistor of 1 ohm is defined by

$$P_X = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)\,dt \tag{1.36}$$

The measurement is performed over some finite time interval $(0, T)$ and resulted in an estimate of the mean power.

The instruments to measure mean power operate on an analog or digital (analog-to-digital) basis. In the analog instruments, the squaring operation is performed via analog squarers built, say, around thermal converters and field-effect transistors.

The variance of a zero-mean random process is obtained by measuring its mean power. The r.m.s. value

$$\hat{U} = \left[\frac{1}{T}\int_0^T x^2(t)\,dt\right]^{1/2}$$

is measured with a square-law electronic voltmeter. The instrument's transducer is remarkable for a parabolic section of the detector characteristic, large bandwidth, and high sensitivity of the amplifier handling the signals of a large peak factor without limiting. For example, the Soviet-made voltmeter B3-48 measures rms values of arbitrary shaped voltage with a peak factor equal to or less than 5 in the frequency range from 20 Hz to 50 MHz. To smooth ripple voltage, the instruments of this type employ an averaging network—a low-pass filter having a very large time constant as compared with the correlation interval of the process under study.

The A/D instruments for measuring mean power and variance are fabricated in three versions.

One version results if the output of the analog squarer is connected to the input of the digital meter of mean value, then

$$\hat{P}_X = \frac{1}{N}\sum_{i=1}^{N} x^2(t_i) \tag{1.37}$$

A second version employs an analog-to-digital time-conversion squarer which performs a conversion of a realization voltage $x(t)$ at the time t_i into a pulse train with the number of pulses proportional to $x^2(t_i)$ to be then averaged on a digital basis.

A third version teams up an analog-to-digital converter working by the successive approximation method, digital multiplier (or digital squarer), and digital averager.

The mean power meters designed to measure unweighted and psophometric power in RF transmission circuits have a built-in unit which automatically presets the duration of measurement. To measure psophometric power, a weighted network (psophometric filter) is used whose weighting characteristic approximately parallels the sensitivity of human ear and modern receiving set. This charac-

teristic is recommended by the CCITT for telephone (voice frequency) channels and by the CCIR for broadcast channels. The psophometric network is usually made as an integral part of the power meter or psophometric voltmeter, and less often as a separate unit.

Measuring correlation functions. Modern correlators make good use of digital techniques along with ordinary analog networks. In addition to routine correlators, the instruments are developed to directly measure the normalized correlation function. The normalized correlation functions are more advantageous than unnormalized in defining normalized power spectra, analyzing slow non-stationary random processes, comparing the stochastic time dependence for analog functions to provide technical or medical diagnosis.

Digital correlators, sampling and quantizing the incoming signal, almost always measure the correlation function in accord with its mathematical definition, that is employing the multiplication technique. The estimates of auto- and cross-correlation functions of ergodic, wide-sense stationary random processes are defined by

$$\hat{K}_X(kT_0) = (N-k)^{-1}\sum_{i=1}^{N-k} x_q(iT_0)\,x_q(iT_0+kT_0) \qquad (1.38)$$

$$\hat{K}_{XY}(kT_0) = (N-k)^{-1}\sum_{i=1}^{N-k} x_q(iT_0)\,y_q(iT_0+kT_0) \qquad (1.39)$$

where $x_q(iT_0)$ and $y_q(iT_0)$ are the quantized values of zero mean realizations of two random processes $X(t)$ and $Y(t)$, n is the number of values measured for the correlation, N is the sample size, and $k = 0, 1, \ldots, n-1$.

Digital correlators utilizing the multiplication technique provide high accuracy of measurement. These devices are built in large part of units and circuits used in computers.

Relatively simple and compact correlators result if they are designed so as to measure the sign correlation functions rather than to operate on the direct scheme. These functions are appropriately broken down into two types, the sign-sign variety and the sign-value variety. Other methods are also in use: sign techniques with additional signals, two-side fast Fourier transform, approximation of the correlation function by an expansion in terms of various basis functions, and that of conditional mean [180].

Measuring the power spectrum. The power spectrum is measured by one of the three methods: filtering, evaluating the power spectral density by the measured correlation function or by taking the Fourier transform of a random process realization (employing the FFT technique).

In filtering, the average power is measured over the known frequency band. If this band Δf is so narrow that the power spectrum

may be deemed constant within it, then

$$F_X(f) \approx P_X(f, \quad \Delta f)/\Delta f, \quad \Delta f \geqslant 0 \tag{1.40}$$

The filter-employing analyzer is made up of the same components as the mean power meter, but in addition the former has a narrow-band band-pass filter, inserted prior to quadratic detector, and a display. In modern analyzers, use is predominantly made of digital filters. To facilitate the analysis, some of the meters perform the time multiplexing of the signal under study.

Evaluating the power spectrum by the correlation function, the measurement is performed through the indirect scheme. The correlation function is measured directly, and an estimate $\hat{F}_X(f)$ is computed as the Fourier transform of the estimate of $K_X(\tau)$. To obtain the consistent estimate of the spectral density, the estimate of the autocorrelation function is multiplied by the weighting function (lag window $h(\tau)$) so that

$$\hat{F}_X(f) = 2\int_0^T h(\tau)\,\hat{K}_X(\tau)\cos 2\pi f\tau\, d\tau \tag{1.41}$$

Analyzers employing the FFT technique have gained wide application nowadays.

Measuring the probability distribution. As will be recalled, the probability distribution function $F_1(x)$ of an ergodic random process $X(t)$ is equal to the relative time a realization $x(t)$ spends below a given level x over the observation interval T as this T increases without bound. For an estimate $\hat{F}_1(x)$, the relative time is used of the realization being lower than the level x for a finite observation period.

One of the simplest in realization is the meter of probability density and function, composed of PAM modulator and mass-produced pulse-amplitude analyzer. Multichannel analyzers equipped with digital discriminators provide great efficiency in measurement and monitoring. When using digital analyzers, some sources of errors can practically be eliminated, such as inaccurate setting of analysis levels, level drift, unequal width of differential quantizing and its drift, and the effect of the frequency responses of the device.

1.6. Some Ways of Further Research

For the years of use, stochastic models of signals, interference and channels have acquired substantial experience in ever widening field of communication problems solved. Yet, there are a large

number of those which are still waiting for their solution. The problem of major importance among them consists in that the adequacy should be ascertained of a statistical model to an actual process for which description this model has been devised. The problem has to be solved whenever some methods of the theory are intended to be applied to practice. The main difficulty here is obviously in that there is no uniform criterion of adequacy and, hence, a common way to verify it for a sufficiently broad class of statistical communication problems of synthesis and analysis so that a constructive, from the practical viewpoint, approach results. Besides, in many applications this adequacy can be verified only experimentally. The respective experiment should be planned on a statistical basis, with the statistical methods of data processing being involved.

As a rule, the stochastic models used cannot exactly describe real processes. Therefore an applicationally important task is to elucidate the stability of the data obtained, subject to variations of statistical parameters of the model. This study is important both in synthesizing a receiver to handle signals in the presence of noise and analyzing the performance of devices with a specified structure under the influence of noise. The problem of stability has been a focus o mathematical statistics in the last years. A possible approach to the synthesis of algorithms under the prior uncertainty is often realized in the development of methods to design and search for the algorithms with various kinds of stability.

As has been noticed, the most economical way to statistically describe interference and signals depends on the structure and type of the device where these signals and interference are observed. Along with the analog devices in which the realizations of signals are continuous functions of time, the digital devices are gaining in importance in communication facilities with the increasingly large proportion in the future, the main advantage of the digital circuits being due to microelectronics. The input and output of these devices are discrete signals. Even though the analysis and synthesis of digital facilities may be worked at drawing on an analogy with the respective analog versions, it is obvious that the techniques should be developed to solve these problems with discrete inputs and outputs. Thus, the need persists to constructively describe discrete signals in terms of probability. So far the designer often has to restrain the study of digital techniques within the framework of the simplest discrete stochastic models. Thus, one of the objectives is to improve the models of discrete signals while keeping the advantage of their simplicity.

For mass-produced and long service-bounded communication apparatus, random deviations in the element ratings may produce an adverse effect on the model adaptability to streamlined manufacture and service life. That these deviations should be mathematically

described in terms of stochastic models is obvious as the cost effect of utilizing an adequate stochastic model may be significant. However, so far the list of such models is far shorter than the list of the stochastic models of signals and noise, which are recommended to solve the problems of reception in the presence of noise.

An extensive work is needed to realize in full extent the possibilities which are opened by statistical design and statistical approach to performance of communication equipment.

Chapter

2

Models of Stochastic System Behavior

2.1. Introduction

The development of science and technology in the second half of this century has among its features the system approach in communication theory and control [420]. This approach consists in that the observed signals are divided into the input and output ones and their relations are described in terms of the characteristic functions and parameters of the processing network. In the simplest case of the linear system the relation between the input and output is given by the convolution operator (the Duhamel integral) which has become a key stone in the classical theory of linear systems.

The methods of system theory have been applied in many fields of communication and control. An expansion of the methods to include random signals was provided by works of Andronov, Vitt and Pontryagin [8], Pugachev [216, 218], Levin [161] and others.

A substantial headway has been made by the advance of the "state variable" technique to give a more general concept of the system [350, 402, 508]. The notion of state played an important part in the development of automata theory. The algebraic methods of the theory were expanded on probability automata. For example, in [402] it was shown that continuous systems and automata might be examined from the common viewpoint and thus a common mathematical theory of deterministic systems might be created. An appreciable contribution to this theory was made by the work of Wunsch [507] and Peschel [452]. The work of Pospelov [214], Bukharaev [39], Kazakov [115], and Krasovsky [152] showed a way to describe stochastic systems by the methods similar to those used in the deterministic theory.

By the time of writing, a great number of works has been published on stochastic system theory and its applications. However, a problem persists to construct models for the objects with stochastic

behavior so that to evaluate what the various approaches and methods of practical analysis have in common. Among the advantages of the unified model the three are especially promising:

The unified model could make the existed mathematical models and methods more comprehensible to the researcher at the common in the engineering quarters intuitive level.

The unified approach to a large variety of problems could facilitate a systematic treatment of system theory for the student.

Finally, the unified technique could enable the result on one discipline to be applied to another, relative to the original in the sense of basic model.

In the following sections of this chapter we shall discuss some aspects of how to build models for the objects displaying a stochastic behavior. In the discussion we shall focus on the ideas and evidence in the model derivation rather than generalizations and mathematical rigour.

2.2. Fundamentals of Deterministic System Theory

The concept of state. The system will be treated as a confined object whose interaction with the surroundings can be represented by the input quantities (cause), that is the influence of the surroundings on the system, and by the output quantities (effect) which depict the reaction of the system to the inputs.

The input and output quantities are functions of time. The behavior of the system is defined by the operator s of the system, which maps the set of all permissible functions X^* on the set of all permissible output functions of time Y^*, that is

$$\text{s}: X^* \to Y^* \quad \text{or otherwise} \quad y(t) = \text{s}[x(t)];$$
$$x(t) \in X^*, \quad y(t) \in Y^* \tag{2.1}$$

The simplest form of s is the direct relation of instantaneous values of the input, x_t, and output, y_t, functions at any moment of time

$$y_t = g(x_t) \quad \text{for all} \quad t \in T \tag{2.2}$$

where T is a set of time points. The systems that obey the equation (2.2) are called *static, instantaneous or memoryless* systems.

In the dynamic systems, the instantaneous value of the output depends on all the values of the input prior to that time[1]:

$$y_t = F(x_\tau; \tau = -\infty, \ldots, t) = F(x_{-\infty}^t); \; t \in T \tag{2.3}$$

[1] We shall denote as x_t^s the segment of the function $x(t)$ on the interval $[t, s)$, as x_{ts} the segment on $[t, s]$, and as x_{t1} the value of $x(t)$ at the time t_1.

If the input process is known only on the finite time interval $[t_0, t]$, then to define the output process $y(t)$ over this interval we require information on the values of the input function prior to the moment t_0, that is on the past of the system. This collection of data about the past of the system, which along with the input specified on the interval $[t_0, t]$ is sufficient to uniquely define the output signal on the same interval, is labelled a *state of the system*. In what follows, the state will be understood as a *minimal* set of such data. The state of a system at a moment t_0 is a link from the past to the future in the sense that it contains all the information about the past that is necessary and sufficient to describe the behavior of the system in the future. It should be emphasized again that the advance of the notion of state and its interpretation has been a significant event in system theory enabling the behavior of various systems to be described from the common positions as will be shown below. The detailed presentation of the general theory of deterministic systems in terms of the concept of state is to be found, for example, in [373, 402, 467, 507, 508].

The description of deterministic system behavior. Let z_{t_0} be the state of the system at the time t_0. We shall describe the system behavior for $t > t_0$ by two equations. The transition equation

$$z_t = \Lambda(z_{t_0}, x_{t_0 t}) \tag{2.4}$$

defines how the system changes its state under the influence of an input signal. The equation of output

$$y_t = \delta(z_t, x_t) \tag{2.5}$$

gives the instantaneous value of the output process as a function of instantaneous values of the state and input process. Substituting Eq. (2.4) into Eq. (2.5) yields

$$y_t = \delta[\Lambda(z_{t_0}, x_{t_0 t}), x_t] = D(z_{t_0}, x_{t_0}^t) \tag{2.6}$$

This equation presents the relation between the input and output, which corresponds to Eq. (2.3) for the case of a finite observation interval $[t_0, t]$. For convenience of practical analysis, we shall transform Eq. (2.4) to the form of the so-called *local transition equation*. For the case of discrete time (T is a set of integers), substituting t for t_0 and $t + 1$ for t gives rise to

$$\begin{aligned} z_{t+1} &= \lambda(z_t, x_t) \\ y_t &= \delta(z_t, x_t) \end{aligned} \tag{2.7}$$

The system of Eq. (2.7) defines the models of synchronous digital automata (when the sets of values of the functions λ and δ are finite or countable) and pulse systems (when the sets of values for the state, input and output are continua).

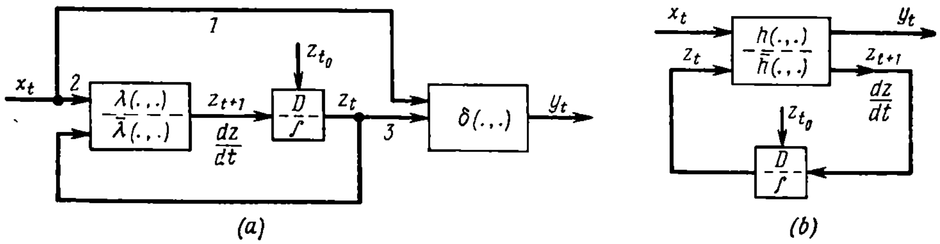

Fig. 2.1. Typical arrangements of dynamic systems

For the case when time varies continuously (T is a set of real numbers) we substitute into Eq. (2.4) $t_0 = t$ and $t = t + h$ and assume that $\Lambda(z_t, x_{t,t+h}) - z_t = 0(h)$. The last implies that with bounded input functions we rule out the stepwise changes of state. Proceeding to the limit as $h \to 0$ yields from

$$dz/dt = \lim_{h\to 0} [(z_{t+h} - z_t)/h] = [\Lambda(z_t, x_{t,t+h}) - z_t]/h$$

and Eq. (2.7) the system of equations

$$\begin{aligned} dz/dt &= \bar{\lambda}(z_t, x_t) \\ y_t &= \delta(z_t, x_t) \end{aligned} \tag{2.8}$$

This model governs the dynamical systems of continuous time. The functional structure of the dynamic systems corresponding to Eqs. (2.7) and (2.8) is depicted in Fig. 2.1*a*. It consists of two instantaneous units which realize the transition function λ ($\bar{\lambda}$) and the output function δ, and a unit delay link, D, for discrete time, or an integrator for continuous time. To note, the structure in Fig. 2.1*a* corresponds to the Mealy automaton (discrete time) or a high-pass filter (continuous time). If the link labelled *1* at the figure is absent, the structure becomes the Moore automaton (discrete time) or low-pass filter (continuous time). Connecting lines *2* and *3* are due to controllability and observability of the system.

The instantaneous units of the state and output can be teamed up into a single static block (Fig. 2.1*b*). Then instead of the system of Eqs. (2.7) and (2.8) we obtain one equation to deal with

$$(z_{t+1}, y_t) = h(z_t, x_t) \tag{2.9}$$

or

$$[dz/dt, y(t)] = h(z_t, x_t) \tag{2.10}$$

A pair (z_t, x_t) is referred to as a situation, a pair (z_{t+1}, y_t) or $(dz/dt, y_t)$ is termed a reaction of the system, and the function h ($\bar{h}$) maps the set of situations on the set of reactions. The configurations *a* and *b* of Fig. 2.1 are in this sense equivalent.

Now we should observe an important property of the transition function in Eq. (2.4). Imagine that the system's state moves first from an instant t_0 to t_1 such that $t_0 < t_1 < t$:

$$z_{t_1} = \Lambda (z_{t_0}, x_{t_0 t_1}) \tag{2.11}$$

and then to a time t:

$$z_t = \Lambda (z_{t_1}, x_{t_1 t}) \tag{2.12}$$

As a result, the system arrives at the same state as in Eq. (2.4) in which case the intermediate state z_{t_1} has not been taken into account. From Eqs. (2.4) and (2.12) with regard to Eq. (2.11) we get

$$\Lambda (z_{t_0}, x_{t_0 t}) = \Lambda [\Lambda (z_{t_0}, x_{t_0 t_1}), x_{t_1 t}], \quad t_0 < t_1 < t \tag{2.13}$$

Equation (2.13) discloses the so-called *semi-group* property of the transition function which is a mathematical expression for the casuality (physical realizability) principle.

2.3. Instantaneous (Static) Stochastic Systems

Introductory remarks. In contrast to the instantaneous deterministic system, the output of the stochastic system cannot be uniquely determined by its input. Because of a stochastic nature of the system, its output, given the input, takes on various values according to some probability law, that is behaves as a random variable. In general, stochastic systems interact with other systems or the surroundings. So it would be natural to treat the input quantity as a random variable, too. Thus, to describe an instantaneous stochastic system one is to specify the relationship between the two random variables. To this end there exist two ways.

Analytical approach (*indirect method*). The objective of this approach is in establishing a relationship between the input and output probability densities. Let us write the joint probability density of input and output random variables, $\xi_t = \xi$ and $\eta_t = \eta$, as

$$w_{\xi\eta} (x, y) = w_{\eta|\xi} (y|x)\, w_\xi (x) \tag{2.14}$$

(as the derived relationships are valid for any t, this subscript is omitted for simplicity).

One-dimensional (marginal) probability density for the output variable η is obtained by integrating $w_{\xi\eta} (x, y)$ with respect to x:

$$w_\eta (y) = \int_{-\infty}^{\infty} w_{\eta|\xi} (y|x)\, w_\xi (x)\, dx \tag{2.15}$$

If η and ξ are discrete random variables assuming the values y_j with the probabilities $p_{\eta j}$ $(j = 1, \ldots, J)$ and values x_i with the probabilities $p_{\xi i}$ $(i = 1, \ldots, I)$, then in place of Eq. (2.15) we get

$$p_{\eta j} = \sum\nolimits_{i=1}^{I} \pi_{ij} p_{\xi i}, \quad j = 1, \ldots, J \tag{2.16}$$

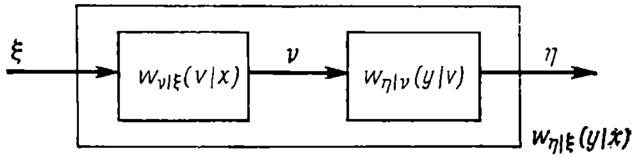

Fig. 2.2. Tandem connected static systems

or in a vector form

$$\mathbf{p}^T = \mathbf{p}_\xi^T \pi \tag{2.16a}$$

where $\pi = (\pi_{ij})_{i=1, \ldots, I;\, j=1, \ldots, J}$ is the condition probability matrix whose elements are $\pi_{ij} = p\,(\eta = y_j \mid \xi = x_i)$, and $\mathbf{p}_\eta$ and $\mathbf{p}_\xi$ are the vectors of the input and output probabilities, respectively. Equations (2.15) and (2.16) may be interpreted as a system relationship of the input-output type in which a linear operator maps the input probability density on the output density. The conditional probability density $w_{\eta|\xi}\,(y \mid x)$ or conditional probability matrix π is the kernel of the operator. Thus we establish a functional relationship between the probability characteristics of input and output variables rather than their values. Thus the name for this method to describe a system—analytical or indirect.

To verify that the function $w_{\eta|\xi}\,(y \mid x)$ is the system characteristic we consider two examples.

(1) *The tandem connection of static stochastic systems.* Let two static systems be specified by the respective characteristics $w_{\nu|\xi}\,(v \mid x)$ and $w_{\eta|\nu}\,(y \mid v)$. We will define the characteristic $w_{\eta|\xi}\,(y \mid x)$ for a system which results from the two tandem connected systems (Fig. 2.2). Recalling the joint probability density rules for the random variables ξ, ν and η yields

$$w_{\eta\nu\xi}\,(y, v, x) = w_{\eta|\nu\xi}(y \mid v, x)\, w_{\nu|\xi}\,(v \mid x)\, w_\xi\,(x) \tag{2.17}$$

Dividing both sides of Eq. (2.17) by $w_\xi\,(x)$ and integrating with respect to v give rise to

$$w_{\eta|\xi}\,(y \mid x) = \int_{-\infty}^{\infty} w_{\eta|\nu\xi}\,(y \mid v, x)\, w_{\nu|\xi}\,(v \mid x)\, dv \tag{2.18}$$

So that to retain the principle according to which we describe a system by a conditional probability density of the output subject to a given stochastic input, we are to obtain in Eq. (2.18)

$$w_{\eta|\nu\xi}(y \mid v, x) = w_{\eta|\nu}\,(y \mid v) \tag{2.19}$$

This implies that the output of the second subsystem of Fig. 2.2, given the input $\nu = v$, is statistically independent of the other variables. Equation (2.19) states what may be called separation condition. (In the deterministic case we shall have instead the prin-

ciple of unumbiguous casuality—the input processes are the sole cause for the output reactions.) Utilizing this condition, from Eq. (2.18) we get

$$w_{\eta|\xi}(y|x) = \int_{-\infty}^{\infty} w_{\eta|\nu}(y|v)\, w_{\nu|\xi}(v|x)\, dv \tag{2.20}$$

Equation (2.20) defines the characteristic of the system by the characteristics of the tandem connected subsystems.

(2) *Defining two-dimensional probability density at the instantaneous system output.* Let at the input of an instantaneous stochastic system be applied a stochastic process $\xi(t)$. In order to define the two-dimensional probability density $w_{\eta_1\eta_2}(y_1, y_2; t_1, t_2)$ of a process $\eta(t)$ observed at the output we shall use the joint probability density of the random variables in the following form

$$\begin{aligned} &w_{\eta_1\eta_2\xi_1\xi_2}(y_1, y_2, x_1, x_2; t_1, t_2) \\ &= w_{\eta_1\eta_2\xi_1\xi_2}(y_1|y_2, x_1, x_2; t_1, t_2)\, w_{\eta_2|\xi_1\xi_2}(y_2|x_1, x_2; t_1, t_2) \\ &\quad \times w_{\xi_1\xi_2}(x_1, x_2; t_1, t_2) \end{aligned} \tag{2.21}$$

As the system is instantaneous (memoryless)[1], we have

$$\left.\begin{aligned} w_{\eta_1|\eta_2\xi_1\xi_2}(y_1|y_2, x_1, x_2; t_1, t) &= w_{\eta|\xi}(y_1|x_1) \\ w_{\eta_2|\xi_1\xi_2}(y_2|x_1, x_2; t_1, t_2) &= w_{\eta|\xi}(y_2|x_2) \end{aligned}\right\} \tag{2.22}$$

Substituting Eq. (2.22) into Eq. (2.21) and integrating with respect to x_1 and x_2 yield the sought expression for the two-dimensional probability density at the system's output:

$$\begin{aligned} w_{\eta_1,\eta_2}(y_1, y_2; t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & w_{\eta|\xi}(y_1|x_1)\, w_{\eta|\xi}(y_2|x_2) \\ & \times w_{\xi_1\xi_2}(x_1, x_2; t_1, t_2)\, dx_1\, dx_2 \end{aligned} \tag{2.23}$$

It can be shown that in the general case the n-dimensional probability density at the system output is defined by the n-dimensional input density and the system characteristic $w_{\eta|\xi}(y \mid x)$ [470].

The probabilistic approach (direct method). For some cases we may succeed in a direct relationship between input and output variables of a stochastic system, namely,

$$\eta = g(\xi, \nu) \tag{2.24}$$

where ν is the independent of ξ random variable which bears the stochastic property of the system. Equation (2.24) may be interpreted as a randomized generalization of Eq. (2.2). As Eq. (2.24)

[1] More general condition for a stochastic system to be memoryless can be stated as $w_{\eta_t|\mathscr{B}_t}(y_t \mid \mathscr{B}_t) = w_{\eta_t|\xi_t}(y_t \mid x_t)$, where $\mathscr{B}_t$ is a set of random variables including ξ_t. The right side in Eq. (2.22) does not depend on t for only stationary systems are considered.

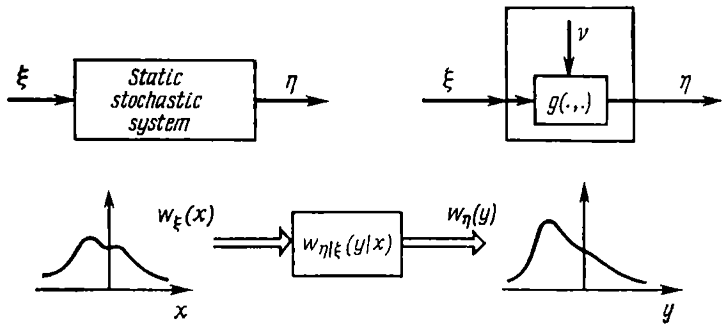

Fig. 2.3. Static stochastic system—direct and indirect descriptions

relates directly the values (realizations) of the input and output variables this method of stochastic system description is referred to as direct or probabilistic. The stochastic system with its direct and indirect descriptions is depicted in Fig. 2.3.

Relation between the direct and indirect descriptions. It is of interest now to answer two questions. May we describe one and the same class of objects by each of the methods? Is it possible to transfer from one method to the other?

To begin with, we will discuss the discrete case when the random variables ξ, η, and ν assume discrete values x_i, y_j, v_k subject to the probabilities $p_{\xi i}$, $p_{\eta j}$, $p_{\nu k}$ ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K$), respectively.

Find an expression for the matrix π [see Eq. (2.16a)] characteristic of the analytical approach. To this end we present conditional probability π_{ij} in the form

$$\begin{aligned}\pi_{ij} &= \mathrm{P}(\eta = y_j \mid \xi = x_i)\\ &= \textstyle\sum_{k=1}^{K} \mathrm{P}(\eta = y_j \mid \xi = x_i,\ \nu = v_k)\\ &\quad\times \mathrm{P}(\nu = v_k \mid \xi = x_i)\end{aligned} \tag{2.25}$$

As ν and ξ are statistically independent, it follows that

$$\begin{aligned}&\mathrm{P}(\nu = v_k \mid \xi = x_i) = \mathrm{P}(\nu = v_k) = p_{\nu k},\\ &k = 1, \ldots, K\end{aligned} \tag{2.26}$$

and as η uniquely (deterministically) depends on ξ and ν [see Eq. (2.24)], it means

$$\mathrm{P}(\eta = y_j \mid \xi = x_i,\ \nu = v_k) = g_{ijk} = \begin{cases} 1, & y_j = g(x_i, v_k)\\ 0, & \text{otherwise}\end{cases} \tag{2.27}$$

Arraying the matrices

$$\mathbf{G}_k = (g_{ijk})_{i=1,\ldots,I;\ j=1,\ldots,J} \tag{2.27a}$$

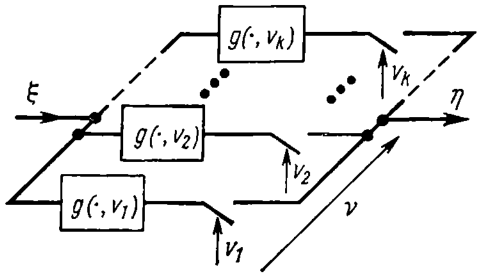

Fig. 2.4. Pictorial representation of Eq. (2.28)

and substituting Eqs. (2.26) and (2.27) into Eq. (2.25) yields

$$\pi = \sum_{k=1}^{K} \mathbf{G}_k p_{\nu k} \tag{2.28}$$

Equation (2.28) shows a manner in which a probabilistic description can be converted to the analytical. As it follows from Eq. (2.27), each value $\nu = v_k$ is supplied by the matrix $\mathbf{G}_k$ that corresponds to a deterministic response $y = g(x, v_k)$. The relation of Eq. (2.28) can be envisaged as a block diagram of Fig. 2.4 in which each randomly appeared value ν switches on its own deterministic system, as it were.

For random variables whose ensembles of possible values are finite, there exists a constructive algorithm to expand the stochastic matrix π in series of the type of Eq. (2.28) (see, e.g. [214]) which proves that in this case a probabilistic description can be obtained from the analytical one. Hence, for the discrete random variables and finite set of values both ways of description are equivalent.

The situation changes as soon as we come over to the systems with continuous input and output variables. In this case the transfer from the probabilistic to the analytical description is effected, with minor constraints, by

$$w_{\eta|\xi}(y|x) = \int_{-\infty}^{\infty} \delta[y - g(x, v)]\, w_{\nu}(v)\, dv \tag{2.29}$$

which is the continuous counterpart of Eq. (2.25). However, it is unknown whether in the general case the indirect description may be back converted to the direct.

Constraints. The general model of the system should be simplified to facilitate the numerical evaluation of the system characteristics in practical analysis. With reference to Fig. 2.5, the model for both additive and multiplicative intraneous interference:

$$\eta = a(\xi) + b(\xi)\,\nu \tag{2.30}$$

If ν is a normalized random variable (of zero mean and unit variance), then we should assume for $a(x)$ the conditional expecta-

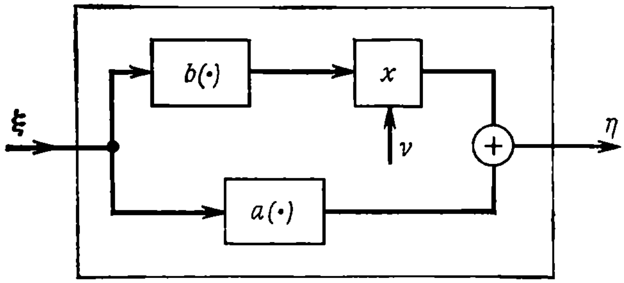

Fig. 2.5. Model of system with additive and multiplicative interference

tion of the output variable at $\xi = x$, and for $b^2(x)$ the conditional variance $\mathrm{E}\{[\eta - a(x)]^2 \mid x\}$.

Note, finally, that the conditions of being separable and memoryless, introduced above for analytical description, are transferred to the required properties of the inner random variable in the probabilistic description. The separability condition [see Eq. (2.19)] corresponds to the statistical independence of the variable ν of all the extraneous disturbing variables (including ξ), whereas the condition of being memoryless [see Eq. (2.22)] implies that the random variables ν_t and ν_s are independent for all $t \neq s$ (a Bernoulli sequence or a random process with independent values at each time point).

2.4. Dynamic Stochastic Systems

State of stochastic system. To simplify the discussion, we shall examine the transitions for the stochastic systems only. Assume that the system under discussion has no output unit (see Fig. 2.1) so that its state is identified by the output process. This turns out to be a way to receiving a fairly broad class of models interesting in numerous applications.

In changing for stochastic models, the notion of state should be sufficiently redesigned. As contracted to the deterministic case, now we should assume that the present state along with the input variate induces a probability distribution of future states rather than defines them exactly. In the lines of the technique used for the static systems, we shall describe these situations by the conditional distribution density, which for the discrete case has the form[1] $w_{\eta_{t+1}\eta_t\xi_t}(y_{t+1} \mid y_t, x_t)$. Multiplying by the joint, for η_t and ξ_t, distribution density $w_{\eta\xi}(y_t, x_t; t)$ and integrating with respect to

[1] Henceforth we shall denote a random process of state (or an output process) as $\eta(t)$, then for a fixed time η_t is a random variable.

x_t and y_t yield the probability density of the state $w_\eta(y;\ t+1)$ in the next moment of time $t+1$:

$$w_\eta(y;\ t+1) = \int_{x_t} \int_{y_t} w_{\eta_{t+1}|\eta_t \xi_t}(y \mid y_t,\ x_t) \times w_{\eta\xi}(y_t,\ x_t;\ t)\, dy_t\, dx_t \tag{2.31}$$

Recurrent representation of the state probability density. Let us reduce Eq. (2.31) to a recurrent form to follow how the state probability density evolves with time. To this purpose we must single out the density $w_\eta(y_t;\ t)$ from the joint density $w_{\eta\xi}(y_t,\ x_t;\ t)$. Examine three cases.

(1) η_t and ξ_t are independent at all t, then

$$w_{\eta\xi}(y_t,\ x_t;\ t) = w_\eta(y_t;\ t)\, w_\xi(x_t;\ t) \tag{2.32}$$

The sufficient condition for these variables to be independent is that ξ_t should not depend on ξ_s at any $t \neq s$ (a Bernoulli sequence).

(2) The input variable is deterministic, that is $\xi_t = a_t$ at all t, then

$$w_{\eta\xi}(y_t,\ x_t;\ t) = w_\eta(y_t;\ t)\, \delta(x_t - a_t) \tag{2.33}$$

(3) The system is autonomous, that is, its input process is a constant [a particular case of Eq. (2.33)], $\xi_t = a$ at all t, then

$$w_{\eta\xi}(y_t,\ x_t;\ t) = w_\eta(y_t;\ t)\delta(x_t - a) \tag{2.34}$$

For these three cases, we can easily obtain the following recurrent formula to compute the state probability density:

$$w_\eta(y;\ t+1) = \int_{y_t} w_{\eta_{t+1}|\eta_t}(y \mid y_t)\, w_\eta(y_t;\ t)\, dy_t \tag{2.35}$$

or in a form of the linear operator

$$w_\eta(y;\ t+1) = \mathbf{T}_t w_\eta(y;\ t) \tag{2.36}$$

whose kernel is the following function for the three cases, respectively:

$$\left.\begin{array}{ll} (1) & \int_{x_t} w_{\eta_{t+1}|\eta_t\xi_t}(y \mid y_t,\ x_t)\, w_\xi(x_t;\ t)\, dx_t \\ (2) & w_{\eta_{t+1}|\eta_t\xi_t}(y \mid y_t,\ a_t) \\ (3) & w_{\eta_{t+1}|\eta_t\xi_t}(y \mid y_t,\ a) \end{array}\right\} \tag{2.37}$$

The recursive relationships of the type of Eq. (2.36) proved very fruitful for the further development of the theory. They help interpreting the state probability density $w_\eta(y;\ t)$ in the sense that a stochastic system in analytical description has a corresponding deterministic system, defined by Eq. (2.36), whose state is the probability density of the state η_t of the original system. It is seen

that in the particular cases discussed the semigroup property of deterministic systems [see Eq. (2.13)] is transferred over the analytical description of stochastic systems. On the other hand, we can prove that this semigroup property of the operator T_t indicates that the random process of the state $\eta(t)$ is Markovian and the expressions of Eq. (2.37) are the transition probability densities and Eq. (2.36) is the Kolmogorov equation [162]. Thus, the stochastic systems of the type discussed are closely connected with Markov processes, the central notion being the concept of state and the central property being the semigroup property of the transition function or operator [321]. In the above discussion we assumed that the input exitations were deterministic processes or processes with independent values. In the next section we shall consider the Markov process input case, too.

Difference and differential equations. To describe a stochastic dynamic system in terms of probability, as we did it in Sec. 2.3, we will proceed along the lines of randomization of the deterministic transition function [compare Eq. (2.2) and Eq. (2.24)]. The stochastic relationship between the random variables η_{t+1}, η_t and ξ_t in analytical approach is described by the conditional probability density

$$w_{\eta_{t+1}|\eta_t\xi_t}(y_{t+1}|y_t, x_t)$$

and with the probabilistic description, is defined by the equation

$$\eta_{t+1} = g(\eta_t, \xi_t, \nu_t) \tag{2.38}$$

with a known probability density $w_\nu(v; t)$ of the random variable ν_t [compare with Eq. (2.7)].

So that the class of models became sufficiently broad, the random variable ν_t should be treated as a random sequence of independent vectors having arbitrary but finite dimension. Thus, the probabilistic description of a discrete-time stochastic system will be in a form of the stochastic difference equation (2.38).

Consider now how Eq. (2.38) transforms in the particular cases of the previous subsection. In case 1 [$\xi(t)$ is a Bernoulli sequence] the pair (ξ_t, ν_t) may be interpreted as another sequence of random independent vectors ν_t^* whose dimension is the sum of the dimensions of ξ_t and ν_t. In case 2 (a deterministic input process $\xi_t = a_t$) the influence of an input variable is in that the function g in Eq. (2.38) will depend upon t, that is, $\eta_{t+1} = g(\eta_t, a_t, \nu_t) = \bar{g}(\eta_t, \nu_t; t)$. In case 3 (the input process is time-invariant: $\xi_t = a$ for all t) the input is absent, as it were, so that $\eta_{t+1} = g(\eta_t, a, \nu_t) = g^*(\eta_t, \nu_t)$. The general equation for all the cases has the form

$$\eta_{t+1} = g(\eta_t, \nu_t; t) \tag{2.39}$$

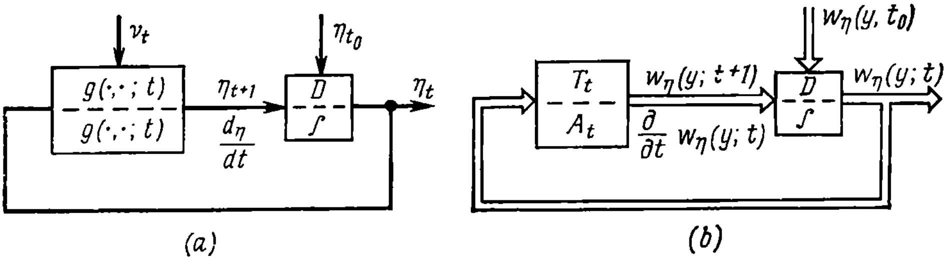

Fig. 2.6. Stochastic system description (*a*) direct, (*b*) indirect

with a Bernoulli sequence of the vectors $\mathbf{v}_t$. The stochastic difference equation (2.39) is solved to yield Markovian processes whose absolute and transition probability densities obey the deterministic equation (2.36).

In the terms used, Eqs. (2.39) and (2.36) are the direct (probabilistic) and indirect (analytical) descriptions of one and the same stochastic discrete-time system.

TABLE 2.1

Set of states	Time	Direct and indirect presentations
Countable (finite)	Discrete	$\mathbf{p}_\eta(t+1)=\pi\mathbf{p}_\eta(t)$ $\eta_{t+1}=g(\eta_t, \nu_t)$
	Continuous	$\frac{d\mathbf{p}_\eta}{dt}=\mathbf{A}\mathbf{p}_\eta(t)$ $\frac{d\eta}{dt}=\bar{g}(\eta_t, \nu_t)$
Uncountable	Discrete	$w_\eta(y; t+1)=\int_x w(y \mid x; t)\, w_\eta(x; t)\, dx$ $\eta_{t+1}=g(\eta_t, \nu_t)$
	Continuous	$\partial w_\eta(y; t)/\partial t=a(y; t)\,\partial w_\eta(y; t)/\partial y + [b^2(y; t)/2]\partial^2 w_\eta(y; t)/\partial y^2$ $\frac{d\eta}{dt}=a(\eta_t; t)+b(\eta_t; t)\,\nu_t$

To transfer to continuous time, we rewrite first Eqs. (2.39) and (2.36) as

$$\eta_{t+h}-\eta_t=g(\eta_t;\ \nu_t;\ t,\ h)-\eta_t \tag{2.40}$$

$$w_\eta(y;\ t+h)=w_\eta(y;\ t)=\mathbf{T}_{t,\ h}w_\eta(y;\ t)-w_\eta(y;\ t) \tag{2.41}$$

Dividing both sides of the equations by h and proceeding to the limit as $h\to 0$, we get from Eq. (2.40) the stochastic differential equation

$$d\eta/dt=\bar{g}(\eta_t,\ \nu_t;\ t) \tag{2.42}$$

(where $d\eta/dt$ should be understood in the sense of generalized random processes) and from Eq. (2.41) the operator differential equation is

$$\partial w_\eta(y;\ t)/\partial t=\mathbf{A}_t w_\eta(y;\ t) \tag{2.43}$$

where $\mathbf{A}_t$ is the infinitesimal operator of the Markov process theory.

Figure 2.6 gives an illustration to Eqs. (2.39) and (2.42) and Eqs. (2.36) and (2.43).

Designation	Application
π = stochastic transition probability matrix g = function of finite (countable) variables with finite (countable) number of values ν_t = Bernoullian sequence with finite number of values	Stochastic automata
$\mathbf{A}$ = transition density matrix g = function of finite (countable) number of values, finite range of η_t and infinite range or relay characteristic with respect to ν_t, (impulse) white noise	Queueing systems
$w(y;\ t+1)$ = conditional transition probability density g = function of uncountable definition domains and values ν_t = Bernoulli sequence of uncountable number of values	Periodic pulse processes passing through dynamic systems
$a(\cdot;\ t)$, $b(\cdot;\ t)$ = continuous functions $\nu(t)$ = normal white noise	Continuous dynamic systems in presence of noise

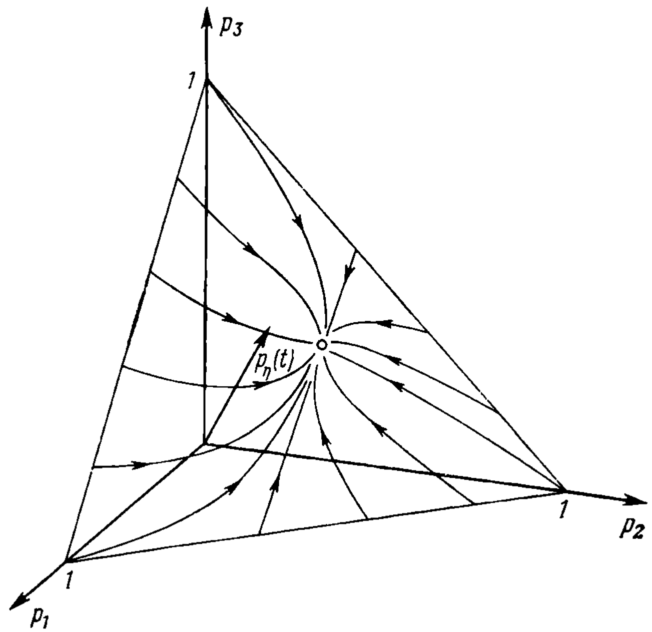

Fig. 2.7. Typical loci of state probability vector

Similarly to the deterministic case (see Fig. 1.2) a dynamic stochastic system may be represented as a static stochastic system connected with a delay circuit and an integrator with a feedback loop.

Functions g in Eq. (2.39) and $\bar{g}$ in Eq. (2.42) and operator $\mathbf{T}_t$ in Eq. (2.36) and $\mathbf{A}_t$ in Eq. (2.43) take on various forms depending on the system. Table 2.1 illustrates the expressions for these operators and functions and clarifies the process $\nu(t)$ in four cases differing in power of both state value and time point sets. Each case is accompanied by an example of typical application.

Geometric representation. We shall add a few remarks on the state space of stochastic systems. To be illustrative, we shall discuss a continuous time system of three states y_1, y_2 and y_3. Then analytically a state will be defined by a state probability vector:

$$\mathbf{p}_\eta(t) = [p_1(t), p_2(t), p_3(t)]^T \tag{2.44}$$

where $p_i(t) = \mathrm{P}(\eta_t = y_i)$, $i = 1, 2, 3$. The manner in which $\mathbf{p}_\eta(t)$ changes with time is described by the equation

$$d\mathbf{p}_\eta/dt = \mathbf{A}\mathbf{p}_\eta \tag{2.45}$$

where $\mathbf{A}$ is the transition density matrix. As

$$\sum_{i=1}^{3} p_i(t) = 1 \quad \text{for all} \quad t \tag{2.46}$$

$\mathbf{p}_\eta(t)$ may move only within the plane defined by Eq. (2.46). Figure 2.7 depicts the space of state with some typical loci of $\mathbf{p}_\eta(t)$. In

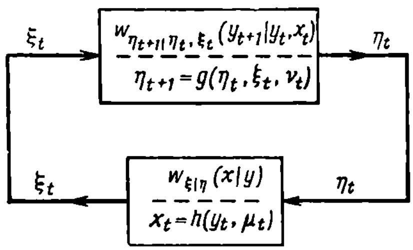

Fig. 2.8. Stochastic system with memoryless feedback

probabilistic description, the states y_i ($i = 1, 2, 3$) may be identified with the unit vectors of the space. Whereas in probabilistic description the states vary jumpwise in the vertices of the simplex at Fig. 2.7, the vector of states in analytical description moves continuously over the face of the simplex.

2.5. Dynamic Systems and Markov Processes

In this section we shall discuss certain applicational examples of the theory, advantages and interrelations between the two methods of stochastic system description.

Stochastic system with instantaneous feedback. With reference to Fig. 2.8, we shall discuss a simple example of a stochastic systems connection. Such systems are a matter of concern with randomized strategy stochastic control theory and learning theory.

To determine the probability density $w_\eta(y;\, t)$ of a process $\eta(t)$ we shall use Eq. (2.31) as a point of departure to factor its joint probability density w_η as

$$w_{\eta_t\xi_t}(y_t,\, x_t;\, t) = w_{\xi_t|\eta_t}(x_t\,|\,y_t)\, w_\eta(y_t;\, t)$$

to obtain

$$w_\eta(y;\, t+1) = \int_{y_t}\int_{x_t} w_{\eta_{t+1}|\eta_t,\,\xi_t}(y\,|\,y_t,\, x_t) \times w_{\xi_t|\eta_t}(x_t\,|\,y_t)\, dx_t w_\eta(y_t;\, t)\, dy_t \tag{2.47}$$

One can verify that from Eq. (2.47) it follows that the process $\eta(t)$ is Markovian [470]. Then Eq. (2.47) is the direct Kolmogorov equation, similar to Eq. (2.36). The kernel of the operator $\mathbf{T}_t$ is uniquely defined by the subsystem characteristics.

With probabilistic description, the Markovian nature of the process $\eta(t)$ is obvious. The system is described by Eq. (2.31) and the feedback by the equation

$$\xi_t = h(\eta_t,\, \mu_t) \tag{2.48}$$

[compare with Eq. (2.24)]. Substituting Eq. (2.48) into Eq. (2.24) we get

$$\eta_{t+1} = g\,[\eta_t h\,(\eta_t,\ \mu_t)\,\nu_t] = g^*\,(\eta_t,\ \mu_t,\ \nu_t) \tag{2.49}$$

Because the random sequences μ_t and ν_t are Bernoullian, the process $\eta\,(t)$ is Markovian (see Sec. 4.2). To arrive finally at the probability characteristics, one is to apply the techniques stated above.

Deterministic systems exposed to random processes. As we have already noticed, the basic model of a stochastic system (see Fig. 2.6) can be interpreted as a dynamic deterministic system exposed to a white noise $\nu\,(t)$. Thus stochastic analysis of dynamic systems is embraced by the basic model. We have assumed so far that input, output, and state processes are scalar variables although all the results are valid in the vector case as well.

A discrete time dynamic system may be described by a difference equation of the mth order. Assuming a Bernoulli sequence as an input signal we obtain the mth order difference equation

$$\eta_{n+m} = f\,(\eta_{n+m-1},\ \eta_{n+m-2},\ \ldots,\ \eta_n;\ \nu_n) \tag{2.50}$$

Knowing m values $\eta_{n+i} = y_i$ $(i = 0,\ \ldots,\ m-1)$, the probability density ν_n and the function f, we can find the probability density η_{n+m}. That means the sequence η_n $(n = 0, 1, 2, \ldots)$ forms a Markov chain of mth order. The same holds true for the solutions of the stochastic differential equation

$$\eta^{(n)} = f\,(\eta^{(n-1)},\ \eta^{(n-2)},\ \ldots,\ \eta;\ \nu) \tag{2.51}$$

[$\eta^{(i)}$ is the ith derivative of the process $\eta\,(t)$], which describes a continuous-time dynamic system. On the other hand, replacing the variables for $\eta_i = \eta^{(n-i)}$, $i = 1, \ldots, n$, yields the equations of state

$$\dot{\eta}_1 = f\,(\eta_1,\ \eta_2,\ \ldots,\ \eta_n;\ \nu)$$

$$\dot{\eta}_i = \eta_{i-1},\quad i = 2,\ \ldots,\ n \tag{2.52}$$

or in the vector form

$$\dot{\boldsymbol{\eta}} = \mathbf{f}\,(\boldsymbol{\eta};\ \nu) \tag{2.53}$$

The last equation is a vector version of Eq. (2.39) and its solution will be an (n-dimensional) Markov process. This exemplifies the known statement in Markov process theory that an n-coupled Markov process may be treated as a component of an n-dimensional first-order Markov process. The reverse is not true [162].

As follows from the above, dynamic system in the presence of white noise can in principle be analyzed within the framework of Markov process theory. However, it does not imply that for every practical case an efficient computing technique exists.

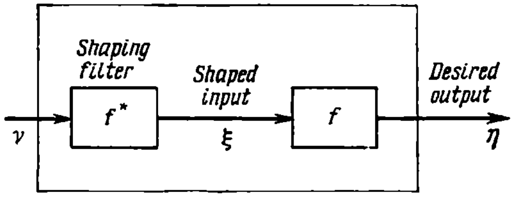

Fig. 2.9. Shaping filter tandem connected to dynamic system

Systems exposed to Markov processes. Let in the stochastic differential equation

$$\eta^{(n)} = f\,(\eta^{(n-1)},\ \eta^{(n-2)},\ \ldots,\ \eta;\ \xi) \tag{2.54}$$

describing a dynamic system the input ξ be an mth order Markov process. Then, at least for a majority of practical cases, there ought to exist a dynamic system defined by the equation

$$\xi^{(m)} = f^*\,(\xi^{(m-1)},\ \xi^{(m-2)},\ \ldots,\ \xi;\ \nu) \tag{2.55}$$

which generates the process from white noise. The problem of practical finding of this system, that is deriving function f^* and distribution ν, is not solved for the general case[1] (compare with the remark in Sec. 2.3 on the statistical description having the analytical one). A standard method to solve this problem exists for a normal input process with rational power spectrum. In this case f^* is a linear function, and ν is a normal white noise.

Suppose that we have succeeded in finding such a system—we shall refer to it as a 'shaping' filter. Then after coming to the state equations for a given system and shaping filter, we obtain [compare with Eq. (2.53)]

$$\begin{aligned} \dot{\boldsymbol{\eta}} &= \mathbf{f}\,(\boldsymbol{\eta},\ \boldsymbol{\xi}) \\ \dot{\boldsymbol{\xi}} &= \mathbf{f}^*\,(\boldsymbol{\xi},\ \boldsymbol{\nu}) \end{aligned} \tag{2.56}$$

Letting η and ξ combine to form the composite vector $\zeta = (\eta,\ \xi)^T$ we can rewrite Eqs. (2.56) as

$$\dot{\boldsymbol{\zeta}} = \tilde{\mathbf{f}}\,(\boldsymbol{\zeta},\ \boldsymbol{\nu}) \tag{2.57}$$

This is the state equation for the two systems being connected in series (Fig. 2.9). From Eq. (2.57) it follows that the vector process $\boldsymbol{\zeta}\,(t)$ is Markovian, thus the processes $\eta\,(t)$ and $\xi\,(t)$ are components of the Markov process. If functions f and f^* in Eqs. (2.54) and (2.55) are such that Eq. (2.57) may be reduced to one differential equation

[1] An approximating approach is presented in [177].

of the type of Eq. (2.51), then we may claim that the output will be an $(n + m)$-coupled Markov process provided that the order of the system is n and a given input process is Markovian m-coupled.

2.6. Concluding Remarks

As we have seen the theory of stochastic processes had acquired a compact form that lends itself to a straightforward analysis. Yet many problems in the field are waiting for their solution. The main of them we should claim the following:

(1) The presented theory is limited with the Markov processes and above all is advanced for the diffusion processes with the normal white noise as exiting process. Though this approach can cover a large amount of applications, the expansion of the theory onto non-Markov processes is, undoubtedly, of theoretical and practical interest.

(2) In bridging from the probabilistic description to the analytical, one has to handle very complicated functional or integro-differential equations. Although a large number of special techniques have been devised already to solve the problem in some particular cases, general efficient computing methods are yet to be developed. Conversion from analytical description to probabilistic in the general case also remains an open problem.

(3) Many fundamental results of random process theory and hence stochastic system theory have been formulated in terms of "pure" mathematics. The engineer without adequate maturity in such subjects as, say, measure theory would meet with difficulties in perceiving and applying these results. Therefore one of the major objectives for engineering faculty members is to present the pertinent fundamental mathematical results in such a way as to make them intelligible to undergraduates, postgraduates, and engineers intended to improve their expertise in the field.

Chapter

3

Coding and Decoding

3.1. Concatenated Coding

Composite codes. It is common knowledge that the efficiency of error-correcting codes is the higher, the larger the length n of these codes [371]. It is only with larger lengths that high fidelity of transmission can be achieved with relatively "poor" channels at a "small" loss of transmission rate. But the longer a code, the more complex are coding, and especially, decoding operations. The complexity of decoding so far has been one of the major obstacles to putting the error-correcting codes into wider practical use.

As the complexity of decoding grows at a much higher rate than the code length, one of the most important objectives, both fundamentally, and especially, practically, is the building of efficient long codes by multiple using of much shorter codes.

The investigation into the problem has led to so-called composite codes where the symbols coded are broken down into various (intersecting) classes, each of which is coded (and decoded after reception) using a sufficiently shorter code. In the process, each information digit is involved in coding and decoding operations with several codes. Moreover, check symbols appearing in coding by one code are then used as information symbols when coding by another (next) code.

The earliest and simplest example of composite codes are Elias's iterative codes [359] in which the configuration of codewords might be described in a two-dimensional form as follows: information symbols (i.e. those to be coded) are presented as an $a \times b$ rectangular table, with the symbols in one row being coded by the same code of length n_2, and after $a \times n_2$ rows are filled out, the symbols in one column are coded with another code of length n_1.

The code of length n_1 used to code columns is referred to as inner, or first level, code; and the code of length n_2 used to code rows the outer, or second level, code. If the distance of the inner code is d_1,

and that of the outer code is d_2, then the distance of the iterated code will be $d = d_1 d_2$.

The representation of codewords of an iterated code as an $n_1 \times n_2$ array suggests a natural way of increasing the distance of a composite code through increasing the number of non-zero columns. In fact, a information symbols in one column will be referred to as a binary representation of an element of the field $GF(2^a)$, i.e. as one symbol of this field; and an ensemble of b symbols will be coded using the outer code over the field $GF(2^a)$ of length n_2. As the basis of external code is increased, its distance may also be increased (until it reaches a limit value $(n_2 - b + 1)$). This modification of the outer code leads to greater distance d of the composite code, which is now estimated as $d \geqslant d_1 d_2$. If the condition

$$2^a \geqslant n_2 - 1 \tag{3.1}$$

is satisfied, then as the outer code one of the modifications of the Reed-Solomon (RS) code may be chosen, for which the distance is equal to the limiting value $d_2 = n_2 - b + 1$. In this case we obtain concatenated Forney codes [365] with the distance

$$d \geqslant (n_2 - b + 1)\, d_1 \tag{3.2}$$

These codes are treated in a considerable detail in works (20-22, 110-112, 397). As the number a is increased further with the condition of Eq. (3.1) met, the encoding and decoding with RS codes becomes increasingly more involved with the distance d_2 remaining the same, then in cases where 2^a exceeds the value $n_2 - 1$ by a factor of two or more, it will be natural to break down the rectangular $a \times b$ matrix into $m \geqslant 1$ horizontal blocks with the number of lines a_i, $i = 1, \ldots, m$, close to $\ln (n_2 - 1)$, but not below it; and to code each of these blocks by its own second level code. Here the resulting distance of the composite code still satisfies the inequality of Eq. (3.2). This approach provides one more way (and rather significant, as will be shown below) of increasing the distance of the composite code.

The fact is that all non-zero codewords of the composite code can be subdivided into m nonintersecting classes such that all the horizontal blocks with numbers $s \geqslant i + 1$ are zeros, and the ith block is non-zero ($i = 1, \ldots, m$).

In this case, for the ith class, in encoding the table $(a_1 + a_2 + \ldots + a_i) \times m$ with (inner) code, codewords of length n_1 may be regarded as codewords of a certain subcode with transmission rate $R_{1i} = (a_1 + a_2 + \ldots + a_i)/n_1$ and distance d_{1i} which is no less than the distance $d_1 = d_{1m}$ of the initial (basic) inner code. In what follows, this subcode will be referred to as the ith inner code or ith code of the first level. The set of all these subcodes ($i = 1, \ldots, m$)

is said to be *imbedded system of inner codes* generated by the basic inner code.

Thus, for the inner codes we have

$$d_{11} \geqslant d_{12} \geqslant \ldots \geqslant d_{1m}, \quad R_{11} < R_{12} < \ldots < R_{1m}$$

As the weight of d_i codewords of the ith class meets the condition $d_i \geqslant (n_2 - b + 1)\, d_{1i}$, it appears that for linear inner and outer codes, i.e. the composite code is linear, its distance will be

$$d \geqslant \min_i \left[(n_2 - b + 1)\, d_{1i}\right] = (n_2 - b + 1)\, d_{1m} \tag{3.3}$$

Hence it follows directly that for the same total number of information symbols of the composite code (i.e. at constant transmission rate) one may, by selecting for each outer code an adequate value of b_i such that $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_m$, make the quantities $d_{1i}d_{2i} = (n_2 - b_i + 1)\, d_{1i}$ be possibly close to each other and above the initial minimum of Eq. (3.3) for all $i = 1, \ldots, m$.

The distance of the composite code $d \geqslant \min\limits_i d_{1i}d_{2i} = \min\limits_i (n_2 - b_i + 1)\, d_{1i}$ may be much higher than the quantity estimated by Eq. (3.3). The composite codes of these types, introduced and studied in (23-25, 106, 107), were termed *generalized concatenated codes* or concatenated codes of mth order. (These include also the codes with the localization of errors.) The first examples indicated that among these there are codes whose distance compares favorably with that of the Bose-Chaudhuri-Hocquenghem (BCH) codes that are some of the best known binary linear algebraic codes.

So, for $m = 2$, $n = 63$, $n_1 = 7$, $n_2 = 9$, $K = 24$ we have $d \geqslant 16$, whereas for BCH codes (63, 24) $d = 15$. At $m = 1$ and the same code parameters we obtain $d \geqslant 12$. For $n = 1\,023$ ($n_1 = 31$, $n_2 = 33$) we get at $m = 2$, $K = 60$, $d \geqslant 384$, and at $m = 3$, $K = 185$, $d \geqslant 240$, while for BCH codes (1 023, 55), $d = 384$ and (1 023, 182) $d = 240$.

Codes with such parameters are not to be found among first-order concatenated codes (Forney codes), and even less so among Elias's iterative codes.

General definition of concatenated code. We would like to introduce some necessary nomenclature. The information word to be coded, μ, will be represented as a rectangular $n_1 \times n_2$ table filled with information symbols located in the first b_i columns of each of m horizontal blocks containing a_i rows each, $i = 1, \ldots, m$. All the other positions of the word μ are zeros.

The $n_1 \times n_2$ array derived by encoding the information digits with external codes (containing zeros in the last $a_{m+1} = n_1 - a_1 - \ldots - a_m$ rows) we shall refer to as auxiliary word and denote by γ. Columns of the auxiliary word will be denoted by $\gamma^{(j)}$, $j = 1, \ldots, n_2$, and the segments of columns containing a_i

digits are denominated by γ_{ij}, $i = 1, \ldots, m + 1$ which are to be treated as the elements of the field $GF(2^{a_i})$ so that $\gamma_{m+1,i} = 0$. Accordingly, each codeword of the concatenated code, i.e. the matrix of the same size, derived from γ by encoding each column $\gamma^{(j)}$ with an inner code, is denoted by α, and its parts analogous to $\gamma^{(j)}$ and γ_{ij} by $\alpha^{(j)}$ and α_{ij}, respectively.

Clearly, in any technique of linear coding with an inner encoder the columns $\gamma^{(j)}$ and γ_{ij} will be related by

$$\mathbf{H}_0^{(j)}\alpha^{(j)} = \gamma^{(j)} \tag{3.4}$$

where $\mathbf{H}_0^{(j)}$ is nonsingular square binary matrix of order n_1.

As the matrix $\mathbf{H}_0^{(j)}$ is nonsingular Eq. (3.4) may be replaced by an equivalent equality

$$\alpha^{(j)} = \mathbf{G}_0^{(j)}\gamma^{(j)} \tag{3.5}$$

where $\mathbf{G}_0^{(j)} = (\mathbf{H}_0^{(j)})^{-1}$.

The matrix $\mathbf{G}_0^{(j)}$, the inverse of $\mathbf{H}_0^{(j)}$, we shall refer to as coding matrix. Should all the columns $\gamma^{(j)}$ be coded with the same inner code, which is always assumed in using the algebraic (non-random) codes, then Eqs. (3.4) and (3.5) can be substituted by the matrix products

$$\mathbf{H}_0\alpha = \gamma \qquad \alpha = \mathbf{G}_0\gamma \tag{3.6}$$

where α and γ are the code and auxiliary words, respectively, which are thought of as binary $n_1 \times n_2$ matrices.

Thus, in the general case, the linear concatenated code of mth order and length $n = n_1, n_2$ is determined by n_2 square nonsingular matrices $\mathbf{H}_0^{(j)}$ (or by matrices $\mathbf{G}_0^{(j)}$) and m linear codes over the field $GF(2^{a_i})$ with b_i information symbols, the word α $(n_1 \times n_2)$ being the codeword of concatenated code if and only if all the horizontal blocks in the word γ

$$\gamma_i = (\gamma_{i1}, \ldots, \gamma_{in_2}), \quad i = 1, \ldots, m,$$

where

$$\gamma^{(j)} = \begin{bmatrix} \gamma_{1j} \\ \vdots \\ \gamma_{m+1,j} \end{bmatrix}, \qquad j = 1, \ldots, n_2$$

are related to $\alpha^{(j)}$ by Eq. (3.4) and are codewords of appropriate outer codes, and the block γ_{m+1} is a zero one.

The difference in matrices $\mathbf{H}_0^{(j)}$ or $\mathbf{H}_0$ (if all $\mathbf{H}_0^{(j)}$ are identical) gives rise to different concatenated codes. So far, the following three concatenated codes have been studied most elaborately:

Type I. The concatenated codes based on the multiplication by a nonsingular square matrix, when matrices $\mathbf{H}_0^{(j)}$ (and hence $\mathbf{G}_0^{(j)}$) are any nonsingular matrices of n_1th order [26].

Type II. The concatenated codes based on the multiplication by a nonsingular lower triangular matrix, when matrices $\mathbf{H}_0^{(j)}$ (and hence $\mathbf{G}_0^{(j)}$) are the lower triangular matrices with all the diagonal elements equal to unity. In a special case, all the diagonal elements of order a_i in these matrices are unity matrices of order a_i (we shall refer to these matrices as special lower triangular matrices [24]).

Type III. The concatenated codes based on the multiplication by a non-zero element of the field $GF(2^{n_1})$. Here, matrices $\mathbf{H}_0^{(j)}$ (and hence $\mathbf{G}_0^{(j)}$) are selected so that the product of Eq. (3.4) is equivalent to the multiplication of a certain non-zero element $h_0^{(j)}$ of the field $GF(2^{n_1})$ by the column $\alpha^{(j)}$ also thought as an element of the field.

Then Eq. (3.4) takes the form $h_0^{(j)}\alpha^{(j)} = \gamma^{(j)}$, where $h_0^{(j)}$, $\alpha^{(j)}$, and $\gamma^{(j)} \in GF(2^{n_1})$. If all $h_0^{(j)}$ are equal ($h_0^{(j)} = h_0$), then $\boldsymbol{\alpha}$ is related to $\boldsymbol{\gamma}$ by $h_0\boldsymbol{\alpha} = \boldsymbol{\gamma}$ or $\boldsymbol{\alpha} = h_0^{-1}\boldsymbol{\gamma}$, where $\boldsymbol{\alpha} = (\alpha^{(1)}, \ldots, \alpha^{(n_2)})$, $\boldsymbol{\gamma} = (\gamma^{(1)}, \ldots, \gamma^{(n_2)})$ are the vectors of length n_2 over the field $GF(2^{n_1})$. The existence of matrices with the mentioned properties for any $h_0^{(j)} \neq 0$, their non-singularity and construction procedure are almost self-evident.

We shall now discuss the main property of the imbedded system of inner codes for the above three types of concatenated codes. That property is crucial for the concatenated codes of order $m > 1$ and is defined by the so-called theorems of embedding [26]. According to these theorems, as $n_1 \to \infty$ there exist nonsingular square matrices, lower triangular matrices including suitable ones to be chosen as matrices $\mathbf{H}_0^{(j)}$, and also non-zero elements $h_0^{(j)}$ of the field $GF(2^{n_1})$, such that all the m inner codes generated by the main inner code have the distance d_{1i}, $i = 1, \ldots, m$, reaching the Gilbert-Varshamov (GV) bound, i.e. $d_{1i} \geqslant n_1\delta_{GV}(R_{1i})$.

From the theorems of embedding at $R_{11} < R_{12} < \ldots < R_{1m}$ it follows that there exists a system of inner codes such that $d_{11} > d_{12} > \ldots > d_{1m}$. Hence, from $d \geqslant d^{(l)} \min\limits_i d_{1i}d_{2i}$, it follows that R_{1i} and R_{2i} can be selected so that for an equal transmission rate R the quantity $d^{(l)}$ grows with increasing order of the concatenated code. Here it is, of course, assumed that for all is the relation $2^{a_i} \geqslant n_2 - 1$ holds. The principle underlying the encoding procedure for the concatenated codes of order $m > 1$ is clear from the above description of these codes. Various encoding techniques are considered in detail in [24, 26].

Decoding. We denote words to be decoded (i.e. error-corrupted words and their elements) and words obtained after the decoding by superscripts "$\wedge$" and "$\sim$", respectively. Then, if we write matrix $\mathbf{H}_0$

in a block form[1]

$$\mathbf{H}_0=\begin{bmatrix}\mathbf{H}_0\\ \vdots\\ \mathbf{H}_{m+1}\end{bmatrix}$$

(where $\mathbf{H}_i$ are rectangular $a_i \times n_2$ matrices, $i = 1, \ldots, m+1$). and recall that $\gamma^{(j)}_{m+1} = 0$ by Eq. (3.4), we shall see that the product $\mathbf{H}_{m+1}\hat{\alpha}^{(j)} = \mathbf{c}_{mj}$ is a syndrome of mth inner code. From syndrome $\mathbf{c}_{mj}$, we decode columns $\hat{\alpha}^{(j)}$ using error-detecting and correcting decoding, i.e. arrive at columns $\tilde{\alpha}^{(j)}$ and construct the vector $\hat{\gamma}_m = (\gamma_{m1}, \ldots, \gamma_{mn_2})$, where $\hat{\gamma}_{mj} = \mathbf{H}_m\tilde{\alpha}^{(j)}$ at those values of j, for which errors were either corrected, or not detected, and $\hat{\gamma}_{mj} = \theta$ is the erasure symbol for those js which errors were detected only. Next, the word $\hat{\gamma}_m$ is decoded by the mth error and erasure correcting outer code, with the result that we get a word $\tilde{\gamma}_m$ or decoding discard (if the combination of erased symbols turns out beyond correction).

In our next step, turning back to columns $\hat{\alpha}^{(j)}$ we calculate the syndrome $\mathbf{c}_{m-1,j}$ of the $(m-1)$th inner code

$$\mathbf{c}_{m-1,j}=\begin{bmatrix}\mathbf{H}_m\\ \mathbf{H}_{m+1}\end{bmatrix}\hat{\alpha}^{(j)}+\begin{bmatrix}\tilde{\gamma}_{mj}\\ 0\end{bmatrix}$$

After the decoding (with errors detected and corrected) we arrive at $\tilde{\alpha}^{(j)}$ (or θ) and construct the vector $\hat{\gamma}_{m-1} = (\hat{\gamma}_{m-1,1}, \ldots, \hat{\gamma}_{m-1,n_2})$, where $\hat{\gamma}_{m-1,j} = \mathbf{H}_{m-1}\alpha^{(j)}$ or θ, that is to be decoded by the $(m-1)$th external code, and so on, until the vector $\hat{\gamma}_1 = (\hat{\gamma}_{11}, \ldots, \hat{\gamma}_{1n_2})$ has been constructed.

For type III concatenated codes, i.e. the codes based on the multiplying by a nonzero element of the field $GF(2^{n_1})$, the procedure of calculating the syndromes $\mathbf{c}_{ij}$ and determining the columns $\hat{\gamma}^{(j)}$ may be connected directly with operations in the field $GF(2^{n_1})$.

It is to be noted that for the type II concatenated codes, i.e. codes based on the multiplying by a triangular nonsingular matrix, coding and decoding procedures may be provided such that with them the information symbols proper, i.e. those to be transmitted, are not the elements γ_{ij}, $i = 1, \ldots, m$; $j = 1, \ldots, b_i$ of the auxiliary word γ, but the elements α_{ij} of the codeword α [24].

The algorithms of concatenated decoding to be recommended for practical use are the following.

Simple algorithm, such that at the $(m+1-i)$th step and ith inner code corrects a predetermined number ν_i of errors and detects

[1] We assume that for all $j = 1, \ldots, n_2$ one and the same matrix $\mathbf{H}_0^{(j)} = \mathbf{H}_0$ is chosen.

all the errors of multiplicity more than ν_i that can be detected; and the ith outer code corrects errors and erasures. With appropriate selection of ν_i the simple decoding algorithm realizes two-thirds of the lower estimate $d^{(l)}$ of the distance. This means that, using the algorithm of concatenated decoding, all the errors with multiplicity $t \leqslant [(2/3)\, d^{(l)} - 1]/2$ are corrected.

Composite algorithm of decoding, such that at the $(m + 1 - i)$th step z simple algorithms are realized with various values of $\nu_1, \nu_2, \ldots$ $\ldots, \nu_z$. From the vectors $\tilde{\gamma}_i$ obtained (not erased) as a result of each decoding operation, we next select only one, following a certain rule. In this case the distance being realized is $[2z/(2z + 1)]\, d^{(l)}$.

It should be noted that in using a composite algorithm of a maximum length such that the numbers ν_s are all possible values from 0 to $[(d_{1i} - 1)/2]$, the distance realizable with concatenated coding coincides with $d^{(l)}$, i.e. the estimate of the lower distance of the concatenated code.

Distance bounds. The distance d is one of the most important characteristics of concatenated codes that is, in a large degree, responsible for their potential correcting capabilities.

Various upper and lower bounds are now determined which characterize the asymptotics (as $n_1 \to \infty$ and $n_2 \to \infty$) of the ratio $d/n = \delta\,(R)$. The lower bounds define the distance attainable with certainty for some classes of concatenated codes, and the upper ones the value which the distance cannot exceed. Thus, in case the upper and lower bounds coincide (for a certain class of concatenated codes), these determine asymptotically the true values of the distance for concatenated codes in this class.

For algebraic concatenated codes the lower estimates $\delta^{(l)}\,(R)$ are given by the relation $d^{(l)} = \min\limits_i d_{1i}d_{2i}$, provided that all the inner codes satisfy the GV bound, and the outer codes are RS codes.

The upper boundaries $\delta^{(u)}\,(R)$ are obtained by estimating from above the maximum weight of special-type codewords belonging to a set thought of as an auxiliary linear code. This auxiliary code is composed of codewords from the concatenated code, which (except for zero ones) correspond to information symbols, containing nonzero symbols only in b_i-th column of ith horizontal block.

Further, the lower bounds $\delta^*\,(R) = d^*/n$ are obtained from the analysis of random concatenated codes, i.e. the codes such that their inner and outer codes are selected at random and independently of a certain set of appropriate codes. The quantity d^* is a distance averaged over an ensemble (determined by the nature of inner and outer codes) of random concatenated codes, and therefore this ensemble contains with certainty a code with $d \geqslant d^*$.

Especially striking results have been obtained from the analysis of lower and upper bounds $\delta\,(R)$ for the concatenated codes of in-

finite order ($m \to \infty$). The internal structure of these codes is decided completely by the function $y = f(x)$, where $x = R_{1i}$, $y = R_{2i}$, provided that for any i we have

$$\lim_{m\to\infty} [a_i/(a_1 + \ldots + a_m)] = 0 \quad \text{and} \quad 2^{a_i} \geqslant n_2 - 1$$

Given below are the basic results of the procedure.

1. For all types of concatenated codes of the structure $y = 1 - \delta_1(R)/\delta_{GV}(x)$ maximizing the lower bound, this bound coincides with the upper one. Thus, we find the class of concatenated codes with the distance $d = n\delta_1(R)$ known within asymptotic accuracy and realizable with concatenated decoding (as $\delta_1(R) = \delta^{(l)}(R)$). Here the true value of $\delta(R) = \delta_1(R)$ at all transmission rates ($0 < R < 1$) appears to be by far lower than the GV bound, i.e. $\delta_1(R) < \delta_{GV}(R)$.

2. For all the three types of random concatenated codes with the structure $y = 1 - 2\delta_2(R)/(1 - x)$ maximizing the upper bound of codes of the second type (i.e. the concatenated codes with inner codes determined by the lower triangular matrices), coincide with the lower order statistic (i.e. averaged over the ensemble) and upper bounds. Thus, one more class of concatenated codes with the distance $d = n\delta_2(R)$ known within asymptotic accuracy, such that for all values of R the relation $\delta_1(R) < \delta_2(R) < \delta_{VG}(R)$ holds.

3. For concatenated codes of the first and third type with the structure $y = 1 - \delta_3(R)\, 2^{1-x}/(2^{1-x} - 1)$, the lower order statistic and upper bounds coincide with each other, which in turn coincide with the GV bound, i.e. $\delta_3(R) = \delta_{GV}(R)$ at all R.

Thus, the third class of concatenated codes is obtained with a distance known within asymptotic accuracy and lying on the GV bound.

It is to be noted, however, that for concatenated codes of structures 2 and 3 (unlike structure 1) the true distance cannot be realized with the above algorithms of concatenated decoding as $\delta_2(R) > \delta_2^{(l)}(R)$ and $\delta_3(R) = \delta_{VG}(R) > \delta_3^{(l)}(R)$. The values of $\delta_1(R)$, $\delta_2(R)$, and $\delta_3(R)$ for various R are listed in Table 3.1.

Practical applications. Distinctive features of the concatenated coding warrant a conclusion as to the possibility of practical applica-

TABLE 3.1

R	0.000	0.007	0.145	0.252	0.500	0.660	0.778	0.859	0.955	1.000
$\delta_1(R)$	0.500	0.350	0.149	0.100	0.040	0.020	0.010	0.005	0.001	0.000
$\delta_2(R)$	0.500	0.443	0.257	0.190	0.093	0.052	0.029	0.016	0.004	0.000
$\delta_3(R)$	0.500	0.451	0.280	0.214	0.110	0.063	0.036	0.020	0.005	0.000

tions of concatenated codes in a variety of channels. To support this we would like to consider three fundamentally different types of channels.

First type—memoryless, low-noise channels, in which only errors with multiplicity smaller than a half of the distance of the code used are to be corrected to achieve a transmission fidelity required (at a given rate).

Second type—memoryless, heavy-noise channels, in which a major portion of errors with multiplicity equal to or more than a half of the distance, is to be corrected.

Third type—channels with memory, which require that the utilized code with a decoding procedure chosen for it should correct errors with sufficiently diverse configurations, e.g. independent errors, single error packets, several error packets of various length, and also a combination of independent errors with error packets.

The first-type channels can successfully use conventional BCH codes with the Peterson-Berlekamp decoding algorithm [325] and codes with threshold decoding [138]. The decoding complexity of these codes has orders $n^2 \log_2 n$ and n^2, respectively. If in the channels of this type, concatenated codes and BCH codes[1] are used as inner codes, and RS codes are used as outer ones, then the order of decoding complexity may be lowered to $n^{3/2}$, thus leading at large n to a marked reduction in apparatus complexity.

In the second-type channels, BCH or other known codes can be used only if the decoding is carried out according to a likelihood maximum algorithms[2], whose complexity is exponential in n. This makes impossible any practical realization of errorless coding for n values anywhere near substantial. At the same time, if the second type channels use concatenated codes with inner codes of length

$$n_1 \approx \log_2 n \tag{3.7}$$

which are maximum likelihood decoded, then the outer channel shaped by these codes becomes, as a rule, a first-type channel. Therefore, as earlier, use may be made of RS code with Peterson-Berlekemp decoding procedure. According to Eq. (3.7), the decoding complexity for the concatenated code has the order of n^2. But unlike the decoding with exponential complexity, this decoding may be realized practically with sufficiently large values of n.

As to the third-type channels, we will, without any attempt at generalization, confine ourselves to one example given in [108].

---

[1] The questions of existence of an imbedded system of BCH codes of length n_1 with transmission rates R_{1i}, $i = 1, \ldots, m$, and the construction of the corresponding matrix $\mathbf{H}_0$ are treated in [25].

[2] Algebraic decoding techniques similar to Peterson-Berlekemp algorithm do not correct all the correctable errors with multiplicity equal to or more than a half of the distance.

This work considers a first-order concatenated code of length $n = 1\,024$ with $K = 456$ information digits and distance $d = 128$ that uses as inner code a binary code with $n_1 = 8$, $K_1 = 7$ and $d_1 = 2$; and as outer code, a lengthened RS code over the field $GF(2^7)$ with $n_2 = 128$, $K_2 = 65$, and $d_2 = 64$. The code in question, with a permanent decoding algorithm, enables errors to be corrected of the following four types: (1) random errors of multiplicity up to $t = 63$; (2) a single error packet of length up to $l = 143$; (3) two error packets of length up to $l_1 = 90$ and $l_2 = 50$ and random errors of multiplicity up to $t = 25$; (4) four error packets of length $l = 50$ each and random errors of multiplicity up to $t = 7$.

There is not a single known conventional (not concatenated) code capable of correcting such a great variety of errors with a permanent decoding algorithm and the same (or close) parameters. It is to be expected that as the order of a concatenated code increases the diversity of types of correctable errors will also grow, the decoding algorithm being the same.

The results provided suggest a possibility of a sufficiently wide practical application of concatenated codes in the designing of communication systems.

3.2. Sequential Decoding

Introductory remarks. As it has been said earlier in the book, as far back as 1949 Shannon indicated that it was possible in principle to transmit data error-free over a communication channel at rates below its capacity. For a long time, however. no practical methods for message coding and decoding were developed, that would allow the utilizing of potentialities of communication channels. A significant breakthrough was the development in 1957 by Wozencraft [505] of a new decoding technique for convolution codes proposed earlier by Elais [360]. Though the Wozencraft algorithm was later rejected, the term invented by him—sequential decoding—now refers to a whole branch of the theory of error-free coding.

In 1963, Fano [361] suggested a new sequential decoding algorithm. It was an improvement over that of Wozencraft and is used in practical work until present, its major advantage being a smaller number of average coding operations required. (The estimate of the average number of operation in the Wozencraft algorithm was false [100].)

Then, in 1966, a sequential decoding algorithm using maximum likelihood approach (ML-algorithm) was worked out [103]. To account for the principal conceptions of sequential decoding we shall consider this very algorithm confining ourselves to binary symmetric channel (BSC).

Maximum likelihood sequential decoding algorithms. The sequential decoding is a method for decoding of the convolution codes. A convolution encoder may be thought of as an apparatus receiving a sequence of information digits, and producing at the output a sequence of coded symbols. In the simplest case, the encoder receives one binary information digit per unit time, and the encoder's output sends to the channel m binary symbols 0 or 1, which is equivalent to the transmission rate

$$R = 1/m \tag{3.8}$$

Each information digit stays in the encoder for a limited predetermined time. Therefore, each information digit can affect only a certain given number (e.g., ν) of code symbols. The quantity ν is said to be the constraint length.

The code sequences at the output of the convolution encoder may be visualized as a code tree. Each branch of the tree corresponds to a certain information sequence, a step upward of a node corresponding to a 0 of the information sequence, and a step downward, to a 1. The symbol sequence written along each branch is a code sequence corresponding to a given information sequence.

The number of code symbols along a branch segment from the initial node to the given one is termed the depth of the node. The presence of constraints length predetermines, beginning with depth ν, the availability of equal segments of code sequences.

When transmitting over BSCs, individual symbols of a code sequence are distorted and the decoder receives a sequence different from the transmitted. The decoder task is to restore the transmitted sequence.

Now we consider an arbitrary node of the code tree. The likelihood function refers to the following quantity

$$z = d\alpha'' + (t - d)\,\alpha' \tag{3.9}$$

where

$$\alpha' = \log_2 2q - B$$

$$\alpha'' = \log_2 2p - B$$

and q and p are the probabilities of true and false reception of a symbol in BSC, respectively; t is the node depth; d is the number of symbols per branch connecting the node with the initial one, these symbols differing from appropriate symbols of an accepted sequence; B is a certain bias subject to

$$R \leqslant B < C$$

The bias B is chosen so that to minimize one or another characteristic of the sequential decoder.

In restoring the sequence transmitted the decoder calculates likelihood functions for various nodes of the code tree. The calculating of likelihood function for a node is called its processing. The principle of the ML-algorithm may be formulated as follows: at any instant of time, the nodes being processed follow directly the node for which the likelihood function is a maximum. We proceed to describe the algorithm step-by-step.

Step 1. Compute likelihood function for the nodes directly after the initial one. Direct their values to on-line storage of the decoder.

Step 2. Arrange all the open nodes being processed in order of decreasing corresponding likelihood functions. (Open node is a node such that at least one of the succeeding nodes is not processed.)

Step 3. Chose a node corresponding to the maximum value of the likelihood function (if several nodes correspond to the maximum value, then select anyone of them). Remove its values from the memory, having preliminarily processed all the nodes immediately following it and having placed the values obtained of the likelihood function in the on-line memory of the decoder.

Step 4. If the decoder first reaches a node at the depth $\tau + (i + 1)\, m$, $i = 1, 2, \ldots$, go over to step 5. Otherwise perform step 2.

Step 5. Make the decision that i-th information digit on the branch where the decoder is located, is true. Remove from the decoder's memory the likelihood functions for the branches on which the ith information symbol differs from the decoded one. Turn to step 2.

The quantity τ is said to be the back search limit. It is easily seen that it is forbidden for the decoder to return back in the code tree by the depth τ and more from the point of maximum penetration into the code tree.

The number of operations spent by the decoder to decode individual symbols is a random quantity, therefore the decoding time is unequal for different symbols. In order to avoid symbol losses in decoding, the sequential decoder is supplied with a buffer, which is normally a displacement register.

The ML-algorithm application technique for other channels is similar to that for BSC, only the likelihood function is defined in each case individually (cf. [104]).

Fano algorithm and algorithm with recurrences. The consideration of the ML-algorithm shows that up to the moment of symbol decoding the decoder stores in its memory all the information that may be required for the decoding. Most of the stored likelihood functions will with high probability not be required for the subsequent decoding. Therefore, some of the likelihood functions may be eliminated from the memory and calculated anew, if need be. This, of course, leads to a somewhat higher average number of decoding operations, but on the other hand, lowers the required capacity of the on-line

memory of the decoder. Both Fano algorithm [371] and algorithm with recurrences [105] are those featuring a low capacity of on-line memory.

The on-line memory of the Fano decoder stores three numbers only: the likelihood function for the node processed; the value of the threshold that takes on only a discrete set of values; and a special variable θ that takes on the values of 0 or 1. The variable shows if the given node is being treated for the first time with the given value of the threshold. The algorithm is formulated so that with a given value of the threshold each node be processed no more than once.

Analogous reasoning underlies the algorithm with recurrences. Here the decoder memory stores likelihood functions for all the nodes of the branch processed, along with the corresponding threshold values. Owing to the fact that in return motion the likelihood functions are not calculated, it is possible to cut the number of operations performed as compared with the Fano algorithm.

Though the recursive algorithm and the Fano algorithm require smaller on-line store, the total capacity of the on-line and buffer storage is larger. Therefore, for decoding in real time domain, the ML-algorithm is preferable.

Characteristics of sequential decoding. When transmitting with convolution code, groups of νR zeros are inserted into the information sequence periodically over n symbol blocks. This, enables first, each block to be decoded independently, and second, error propagation to be avoided. The remarkable characteristics of the decoder are: the probability of false decoding of a block (when at least one symbol of the block differs from the transmitted); the average number of decoding operations per block; the probability of overflowing of the buffer and on-line storage of the decoder.

Experimental findings and estimations indicate that the probability P (E) to erroneously decode a block of n symbols drops exponentially with increasing constraint length

$$\mathrm{P}(E) = ne^{-[E+o(\nu)]\nu} \tag{3.10}$$

where the reliability function, E, varies with the code, bias B, transmission rate R, and ratio τ/ν; here $o(\nu) \to 0$ as $\nu \to \infty$.

The average number of operations per block N at $B = R$ and transmission rates R below the channel capacity,

$$R_{comp} = 1 - \log_2 (1 + \sqrt{4pq}) \tag{3.11}$$

remains an unlimited quantity, independent of ν. We note that among the three algorithms considered, it is the ML-algorithm that has the minimum number of operations. The second is the recursive algorithm. The probability that the buffer (on-line) storage of the decoder is overflown is a power function of the memory capacity L,

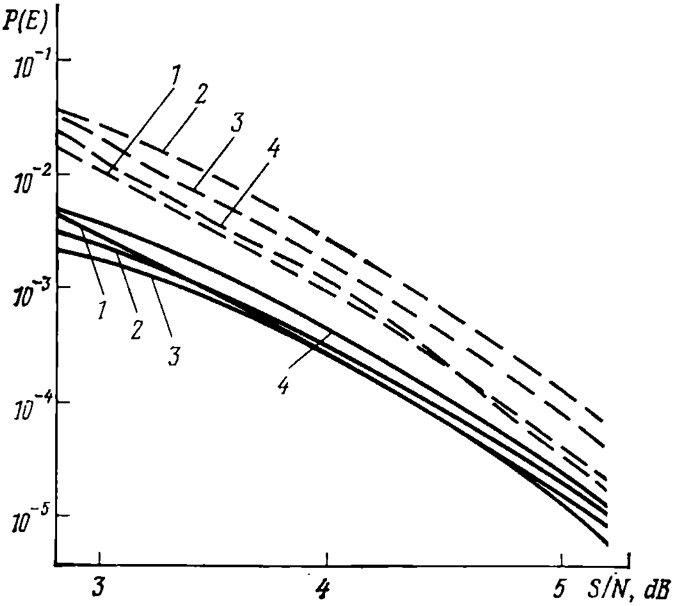

Fig. 3.1. Probabilities of false decoding (—) and erasure (— — —) of a block: *1*—Viterbi decoder; *2*, *3*, *4*—ML-decoder with on-line store capacity 600, 800, and 1000 nodes, respectively

that decreases as $L^{-\beta}$, where β depends on the transmission rate and channel parameters.

In actual practice, the sequential decoders are normally used at transmission rates smaller than the computational cutoff (usually $R = 0.9R_{comp}$). The magnitude of ν is selected so that the probability P (E) of false decoding would take on values 10^{-5}-10^{-7}; the buffer and on-line storage capacities are chosen such that the probability of their overflow would not exceed 10^{-3} to 10^{-4}. In recent years, the sequential decoding is utilized at transmission rates higher than the calculated channel capacity for reasonable values of the parameter ν. In this case, the sequential decoding algorithms compare favourably with the Viterbi decoding algorithm.

Tradeoff of decoding methods. First we compare sequential decoding with Viterbi's decoding of convolution codes. The Viterbi decoding technique compares completely the code tree branches with the received sequence adopted. It uses in decoding all the information contained in the coming signal, although at the expense of high computational requirements.

Figure 3.1 compares the results of the ML-decoder with that of the Viterbi decoder carried out in the Institute for the Problems of Information Transmission of the USSR Academy of Sciences. The decoding was considered for a convolution code with shift register

length 8 and transmission rate $R = 2/3$ (at code constraint length $v = 12$) for the "hard" decision (i.e. for the case of binary input and output alphabets) for the cross-over probability, in the channel ranging from 0.025 to 0.005, which corresponds to signal-to-noise ratio from 2.8 to 5.2 dB. The memory of sequential ML-decoders was from 600 to 1000 nodes. Both the ML- and Viterbi decoder allowed for symbol erasure. The transmission occurred in blocks with $n = 48$ information digits apiece, appended by six zeros.

The experiment evidences that even with relatively short constraint length the sequential decoding compares favourably with Viterbi's decoding, and even surpasses it by a number of parameters. In particular, in experiments with an EC-1030 computer the average decoding time for a block ranged from 0.5 s ($p = 0.005$) to 2.5 s ($p = 0.03$), whereas for the Viterbi decoder it took 12 s to decode.

Similar conclusions were drawn in paper [339], devoted to the development of a form of ML-algorithm called by the authors the "complete" algorithm of sequential decoding. This algorithm differs from the above ML-algorithm in that it lacks decoding failures caused by on-line memory overflow, and in case the memory is overflown a part of considered branches is rejected positively with the decision made then concerning the left branches only. A case was investigated where the transmission rate was $R = 1/2$ bit per symbol, and the length of the shift register took on the values 3, 4, 5, 7, 12, and 15. The experiment has shown that at equal reliability the sequential decoding requires several times fewer operations than the Viterbi algorithm. So, to achieve an error probability of 10^{-5} at an S/N ratio of 5.5 dB, the Viterbi algorithm takes by two orders of magnitude more operations than the sequential decoding.

The results of the tradeoff of various algorithms are given in Table 3.2 that lists shift register lengths m, the average number of operations C and the required capacity of the on-line storage. The

TABLE 3.2

Error rate per bit	Viterbi algorithm (hard decision)			ML-algorithms (hard decision)			Viterbi algorithms (8-level quantization)		
	m	C	memory, K	m	C	memory, K	m	C	memory, K
1.5×10^{-3}	3	504	16	5	201.0	400			
5.0×10^{-4}	4	1024	32	5	81.9	530			
7.0×10^{-5}	7	8576	256	12	82.3	1640			
3.5×10^{-5}			12	12	82.9	2400	2	248	8
1.5×10^{-5}				15	84.7	3700	3	504	16
7.0×10^{-6}				15	85.1	5000	4	1024	32

signal-to-noise ratio is 5.5 dB, the number of symbols per block is $n = 60$. The last column in the table refers to the so-called "soft" decision. Most of actual channels reduce in theoretical treatment to channels with continuous additive noise.

In decoding, the incoming signal is quantized into several levels. If the number of levels is two, we have a decoder with hard decision; if more, the decoder with soft decision. It is seen in the table that as the number of quantization levels is increased, the noise immunity of the system is markedly improved. Unfortunately, the data as to the use of sequential decoding with soft decision, are not available so far.

The investigation into the theoretical and experimental evidence warrants a conclusion that the Viterbi decoder is the better choise with relatively good channels, and/or moderate requirements on the output probability of error. For poor channels and with stringent requirements posed on the decoding reliability it would be advisable to use the sequential decoding or concatenated codes.

Practical applications. The sequential decoding is a relatively involved and expensive form of code protection in communication channels, therefore its use is only justified in special-purpose channels, e.g. in some satellite communication systems. The realization of the sequential decoding apparatus requires several hundreds of large-scale integrated circuits (LSI). With rather low data transmission rates (several thousands of bit/s) for decoding use may be made of a medium-size computer.

The sequential decoding is a good practice, first of all, in channels with independent errors. These channels are known to include satellite channels. Reference [432], for instance, provides the results of utilization of sequential decoding in a spacecraft communication system. Reportedly, the sequential decoding can also be used in channels with small error packets [201]. Longer packets in communication channels sharply increase the number of operations, and hence result in the memory being overflown. To combat the error packets the symbols are often interlaced, or retransmitted on request via a feedback channel.

3.3. Coding in Channels with Random Structure

State-of-the-art. Many important problems of error-free coding which are waiting for exact solution include the problem of coding with messages transmitted over random structure channels (RSC). The name normally refers to the channels which apart from random additive noise, are characterized by random variation of their parameters, e.g. random parameters of the transfer function (see Sec. 1.4.). In using various modems the continuous channels are mapped into rather complicated discrete channels whose parameters

can also be described by certain random sequences, continuous or discrete. Unlike the simplest continuous channels with additive normal white noise which in linear modems more often than not are mapped on a discrete symmetric memoryless channel (DSC), stochastic continuous channels are for the most part mapped onto discrete channels with memory, not always symmetric at that.

Random structure channels are widely utilized in communications. These include essentially all the radio channels, safe may be for those which provide the line of sight communication between the transmitting and receiving antennas, if the path does not involve reflectors and absorbers moving about at random. In most cases switched cable channels also feature a randomly varying structure. The list of examples can be increased by hydroacoustic channels, and many optical communication channels.

For a long time the theory of error-free coding centered mainly on DSC with independent errors. This resulted in the fact that many codes developed during the 50s and in the early 60s failed to find wide use in RSCs, at any rate in systems with directly corrected errors (without any feedback). The earliest, rather primitive, noise-immune codes to gain recognition in practice came into being, so to speak, away from the main path of development of coding theory. These took into consideration the significant difference of an actual discrete channel from DSC. Worth mention here is the Verdan code used in the Baudot radio system and based on a simple repetition of a code combination. The error-correction took account of the marked asymmetry of the discrete channel resulting from on-off keying in a radio channel with fading and concentrated interference. This system was employed in the 1930s when the problem of noise-immune coding has not yet been formulated.

Another, much later example is the utilization of a permanent-weight code (3-out-of-7) for radiotelegraphy with feedback [358]. Here also the coding took into account the channels asymmetry which makes the probability of the error, that does not affect the weight ("shift" error), negligibly small, whereas all the remaining errors are detected by the code.

In the mid-1950s works began to appear which were devoted to optimal processing of signals in a continuous channel. These treatments considered, along with additive noise, the random variations of channel parameters in the form of slow fluctuations of the phase, and later the amplitude, too. By that time the channel with slow Rayleigh fadings has been investigated in a considerable detail, and somewhat later papers began to be published concerning channels with multipath propagation of the signal (see Sec. 1.4, and also [267, 455, 492]). Even for these, relatively simple RSC models the discrete mapping, as a rule, differs substantially from DSC. In particular, when transmitting discrete messages over a channel

with slow fading errors tend to group together. But the works devoted to the optimal reception of signals in continuous RSC overlooked the problem of coding.

Until fairly recently the discipline of communication theory was characterized by strict division of spheres with some workers dealing with the optimal signal reception in the continuous channel, and the others with the coding in the discrete channel. The former sought the realization of an algorithm for signal processing such that would minimize the loss function. Most of the treatments considered the element-by-element processing only, with the unconditional error probability being taken to be the loss function. On the other hand, the latter concerned themselves with discrete channels regardless of corresponding continuous channels, constructed various mathematical models of the channels and sought coding techniques suitable for these models [19].

So far it is widely believed that if a code is good for a DSC, it is bound to be quite satisfactory for any other binary channel as well, though to ensure the specified fidelity the length of the code block (or constraint) should be somewhat increased. This stand is normally supported by the fact that the capacity of continuous channels with random structure is, as a rule, but slightly different from the channel with normal additive white noise. But in principle, of course, any correcting code might be utilized in a channel with random structure, providing that the constraint length is sufficiently large for the number of errors in a block to obey the law of large numbers. This approach, however, results in unrealizable systems. So, in a short-wave radio channel fadings are rather common with a period of the order of several seconds. With conventional element-by-element transmission, e.g. FSK at a reasonable rate of 100 bit/s, in order to efficiently average the error multiplicity blocks of up to 10^4 digits have to be used. In data transmission over switched cable channels at 2 000 bit/s, according to CCITT data, in about 10% cases there occur interruptions of length over 3 s, i.e. involving up to 6 000 digits. An error correction with a code designed for DSK would require a distance of over 12 000, i.e. the length of the block of the order of 10^5 digits and more.

One of the earlier ideas in the coding of discrete channels with dependent errors was to separate in time the digits used in the general check, the idea being suggested first for recursive codes [379] and later for block codes [18]. In this way the errors are decorrelated, i.e. a channel with errors grouped into packets turns into a channel similar to a DSK, which allows the using, say, 1 000 intermittent 100 digit blocks instead of one 10^5 digit block. This of course, obviates the realization difficulties, as larger storage capacities are required, but the decoding is materially simplified and with the present state-of-the-art is readily achievable.

But this coding technique is not optimal for channels with grouping error, as it does not use the information on error correlation. In particular, in decoding the separated blocks, no information is used pertaining to the results of decoding of symbols closely spaced in time but belonging to other blocks. In recent years methods have been developed which enable this information loss to be avoided [147]. In addition, codes, became available which detect and correct error packets [309]. The correcting of error packets did not find wide recognition in practice, mainly owing to complexity and insufficient effectiveness of appropriate algorithms.

Linear stochastic channels. At present, linear channels with random structure, or linear stochastic channels (LSC), receive primary attention which is in full accordance with practical demands, as the overwhelming majority of the channels used lends themselves to description by this model. The only exception are the digital channels with regenerators that are substantially nonlinear. References [128, 404] consider the continuous LSC regardless of the coding in the conventional (discrete) sence. Also, these texts deal with the question of the optimal signal selection, their optimal processing, and suggest mathematical models of LSC. The classification of LSC models is given in [147].

All LSCs may be described with the aid of a random transfer function depending, in the general case, on two arguments. In special cases, LSCs are described in a much simpler way. From the random transfer function we may, using bilateral Fourier transforms, obtain a number of "system" functions characterizing LSC from various angles [128].

As signals pass through LSC they normally exhibit the spreading of spectrum and extention of it in time. In special cases one of these phenomena prevails.

The principal task in design of discrete message transmission systems consists in developing the apparatus connecting the given continuous channel with the message source and destination. In principle, the task is solved by comparing each of possible messages being transmitted with one of the signals sent into the channel. The maximum fidelity of reception is provided, if the decision on the received signal is made on the maximum likelihood basis, and the very signals are selected considering the channel characteristics so as to be best identified by a receiver. Under certain conditions, the false reception probability tends to zero, if the duration of the message compared with each signal increases. In the process, the number of signals grows exponentially and such a transmission technique becomes impossible to implement. Therefore, fixed constraints are introduced, such that for the channel the input signal at any instant of time depends on no more than a predetermined number of message symbols. The coding problem consists in devising regular (non-

exhaustion) methods of correlation of the signals with sequences of symbols of the source, and the decoding problem consists in implementing the reverse procedure.

More often than not, these procedures occur in two stages: the coding-decoding proper, and the modulation-demodulation. The modulation turns the continuous channel into semi-continuous with discrete symbols fed to the input, and continuous signals obtained at the output (normally as readings in the computing circuit that are the functions of discrete time). After the decision is made the semi-continuous channel in the demodulator is transformed into a discrete one.

At the present time, only those modulation and demodulation techniques are utilized which might be reduced to linear operations with a sequence of discrete symbols and certain basic functions. The properties of semi-continuous and discrete mappings obtained from LSC for "linear" modems are discussed in [147]. Based on the results of the examination of these mappings the problem may be solved of the selection of coding methods for RSC with decoding in the discrete channel (usually an elementwise reception), or in the semi-continuous channel (reception "as a whole" and various suboptimal forms of the analog decoding). Note that the elementwise reception can only be optimal on the condition that the semi-continuous channel is memoryless.

An M-ary channel is regarded as specified if for any n we know the conditional probability distribution $P(\hat{\mathbf{a}}|\mathbf{a})$ for the arrival of the sequence $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_n)$ of M-ary symbols of length n, once the channel input is the sequence $\mathbf{a} = (a_1, \ldots, a_n)$. A sequence consisting of 0s and 1s, in which 1s occupy bits where $\hat{\mathbf{a}}$ and $\mathbf{a}$ do not coincide are named the error pattern. A discrete channel with additive noise refers to a channel in which $P(\hat{\mathbf{a}}|\mathbf{a})$ is only dependent on the error pattern and independent of the sequence transmitted.

Non-additive noise in a discrete channel is, as a rule, bound up with asymmetry of the channel or with the intersymbol interference. Many a discrete channel with additive noise may be described using a model of a channel with variable parameter (CVP) where the transition probabilities for symbols are conditioned by a certain (generally, vector) random process.

It is worth noting that the characteristics of the discrete mapping depend not only on the characteristics of the continuous channel, but on the modem used as well. This, unfortunately, is often overlooked. So, speaking about error grouping, for instance, in a shortwave or radio relay channel, no reference whatsoever is made to the modulation type. Hence a new objective of modem optimization: instead of seeking the minimum for the error-reception probability of a symbol, the criterion should be used for the minimum for the false decoding probability of a message. Therefore, in the communi-

cation system design the joint optimization of the modem and codec may be regarded as fairly promising.

In [147] basic data are given for some forms of CVP, which are needed in working out the coding procedures. In particular, it provides the probability of errors of given multiplicity, and also the probabilities of certain error patterns. Moreover, bounds are obtained for achievable probabilities of true decoding of a block. Also of importance are the results of this work pertaining to the dependences and bounds for the probabilities of errors undetectable by code in systems with feedback.

Coding and decoding. Unlike DSC that is completely characterized by the one parameter—probability of error, the discrete mappings of actual RSCs are rather varied and their description takes several parameters. A two-parametric mathematical model of discrete channel (see [219]) allows only a rather rough approximation. The variety of discrete channels gives rise to one more aspect of the problem of error-free coding, viz., the matching of coder and channel parameter. Obviously, it is practically impossible to devise a special code for each channel, even more so because the channel cannot be thought of stationary for long spaces of time, i.e. the possibility of its parameter variation should be taken into account.

Two avenues may be followed in the design of coding system suitable for the channel of every description, adaptation and versatility. The first one is to measure channel parameters before the coding (or before the decoding only), and to change the algorithm used accordingly. The channel parameter measurement ("teaching") is carried out, for instance, using a special test signal or from the results of reception of preceding segments of the information-carrying signal, or from a preliminary analysis of the signal segment to be decoded [73, 132, 133].

The other approach relies on the fact that under certain circumstances coding techniques may be suggested, which are optimal or near optimal for a broad class of channels. Furthermore, to transform the channels we may devise methods allowing us to reduce various initial channels to a certain "standard" discrete channel, for which the optimal coding is known. A simple example of this approach is the above-mentioned method of error decorrelation by separation of symbols belonging to one code block. This method permits of approximate transformation of a great variety of discrete channels into DSC. On the whole, the adaptive methods are bound to provide higher noise-immunity than universal ones. But in most cases the universalization enables simpler circuits to be constructed.

The decoding universality problem gives rise to the question of whether it is possible to utilize a simple Hamming algorithm of decoding by minimum distance. In DSC this algorithm corresponds

to the maximum likelihood criterion, as all the error patterns with equal weight are equiprobable and their probability drops monotonously with weight. In discrete LSC mappings, even with additive noise, the probability of an error pattern is not uniquely determined by weight and is not bound to decrease as the latter increases. Nevertheless, not infrequently the code selection by the maximum of minimum distance and the decoding using the Hamming algorithm are thought of to be close to optimum [213].

It is argued at times that the Hamming algorithm is suitable for CVP (that is called the composite algorithm [503]), if the parameter varies so slowly that during the time as the block passes it may be regarded as constant. In fact, in this channel during the passage of a block the error probability is fixed, and all errors are independent. But blocks received with small error probability do not use the redundancy of the code, whereas in blocks received with higher error probability the redundancy appears insufficient and the errors remain practically uncorrected.

Note that with the redundant coding in a composite channel not for the direct correction of errors, but for their detection, the conditions appear to be absolutely different. The probability that the block falsely received in channel dropout will be corrected, is negligible, whereas the probability that in this case the error will be detected is close to unity.

So far there are not many constructive algorithms of coding and decoding with error correction in the discrete channel with memory, even in such a simple channel as CVP (without making resort to a continuous channel and feedback). Besides the above-mentioned code with error packet correction and codes with alternating blocks used to decorrelate errors, mention may be made of but some methods of concatenated coding used mainly for "vector" channels, as well as a form of the method of alternating blocks which considers the correlation between errors in "neighbouring" (in time and frequency) symbols of various blocks. The last method is rather promising, especially so for the channels with notable correlation of errors. The alternating enables the symbols with practically independent errors to be combined into blocks. After a first block has been decoded, symbols are known that have been received with errors. This suggests that the neighbouring, second-block symbols, which are rigidly correlated with them, are unreliable. A knowledge of this allows us to decode more reliably a second block, and so on. Still more effective results are obtained using an analogous method for a concatenated code [147]. We shall also mention here the utilization of convolution codes with sequential decoding following a generalized metric taking into consideration the properties of discrete LSC mapping, and also the use of simple recurrent codes [140, 340, 341].

New possibilities are offered by the use of decoding in a semi-continuous channel. Most of the known techniques use the signal at the output of the semi-continuous channel to evaluate the reliability of various components of discrete channel and to derive weighting coefficients for all sorts of test. Although for the most part these methods have been developed for memoryless channels, they turned out to be the most convenient in CVP decoding [32, 189].

It is to be noted that the analysis of analog methods of decoding in a "conventional" channel with additive normal noise leads to rather pessimistic conclusions, as the gain in energy here as compared with the more simple discrete decoding appears to be negligible [506]. In RSC, however, these methods give a by far more marked effect.

We know many decoding methods for continuous or semi-continuous channel, from the reception "as a whole" where the demodulation and decoding are completely combined [189], to the algorithms where the continuous signal is only utilized to establish the rank statistics, i.e. to arrange the symbols obtained in the demodulation in decreasing order of reliability (e.g., the method of decoding by the most reliable symbols [32]). An intermediate position is occupied by the algorithm in which the signal at the output of a semi-continuous channel is utilized to derive the weighting coefficients [114]. Mention should be here made of works [337, 338] dealing with the rank algorithm of analog decoding and giving the results of its modeling and testing in actual channels.

Any further improvement in conditions of discrete message transmission over RSCs is only possible with a new approach to the combined development of the modem and codec. As pointed out above, the modem determines to a large measure the characteristics of discrete mapping. Almost for any LSC a modem can be designed for example, such that a discrete mapping close to DSC could be obtained. But this is not always the optimal method of using the modem, and is often conclusive to unsatisfactory results as to transmission rate and error probability.

The joint optimization of the modem and codec is not sufficiently studied so far. We shall dwell, therefore, on multi-dimensional or vector discrete channel. The simplest example is the two-dimensional frequency-time channel using multi-frequency modems. In this case, the symbols to be transmitted and received are arrayed as a matrix in which the rows correspond to various time intervals, and the columns to various subcarrier frequencies. A distinctive feature of such channels is the presence of correlation intervals along the frequency and time axes. Also, it is possible, by varying the modem parameters, and maintaining the same information transmission rate, to redistribute memory in the discrete channel, for instance, to enhance the correlation between symbols occurring close in time

and reducing it between symbols closely spaced in frequency, and vice versa. So, for instance, an inter code may be utilized along the coordinate with "long" memory and decoded only when errors are detected and corresponded sub-blocks are erased. Symbols of the outer code appear to be but weakly correlated with the result that erasures and a certain number of unerased errors may be corrected effectively and relatively easily. Modems making up a two-dimensional channel will be also useful for such continuous channels in which there are heavy concentrated interferences and impulse noise. Here errors show a tendency to mutilate the symbols transmitted on one subcarrier frequency (with concentrated interference), or within one time interval (impulse noise). Coding methods are now available that are also successful in correcting the lattice-type error patterns [147].

Transmission of information over RSC with feedback. Essentially all of the data communication systems now in operation use the computing feedback. As it was indicated above, the coding and decoding with error detection is implemented in CVP more effectively than the error correction proper. One of the reasons of it is the fact that the optimal decoding algorithm with error detection is absolutely independent of the channel characteristics, and the probability of error being undetected is but slightly dependent on the channel characteristics, but is conditioned mainly by the code structure.

It has been proved that for any n and k a code is available that detects errors uniformly, i.e. provides in any DSC a probability of error being undetected of no more than $2^{-(n-k)}$. This is also valid for any composite CVP. Now we have also methods for universal stochastic coding providing this boundary for the probability of error being undetected in any binary channel [146, 148]. Note one rather substantial but obscure point. If two discrete channels were compared, one memoryless (e.g., DSC), and the other with a distinct memory (e.g., composite), the unconditional errors probability, then the system with a computing feedback will ensure a much better information transmission with the second (composite) channel.

In DSK the number of errors in each block varies within narrow bounds, but blocks received at least with one error are requested for repetition. Therefore, the number of errors in a discarded block is immaterial.

The composite channel will have many channels with a large number of errors, but on the other hand, many blocks received without error, and hence less repetition requests. The information transmission rate over the composite channel will thus be higher than over an equivalent DSC. Further, the residual probability of false decoding is also dependent on the probability of request for repetition, and in the composite channel it is less than in DSC.

3.4. Efficiency of Communications with Correcting Codes

Efficiency criteria. The most important parameters of a system transmitting discrete messages are the transmission rate, R bit/s, and the probability of error, p, characterizing the fidelity of the transmission. The combination of these two variables defines the effectiveness of the system.

In systems without coding error probability may be decreased only at the expence of a higher signal power or lower transmission rate. With stringent requirements on the transmission fidelity it becomes desirable to utilize correcting codes (at first more simple, block or convolution codes with threshold decoding). In cases where both high fidelity and high rate of the data transmission are required, use is made of more powerful (and hence more complex) codecs.

Now we consider a system where each of $M = 2^k = 2^{RT}$ signals is represented by a sequence of n binary symbols (the number of such sequences is 2^n). The following estimate of the average error probability may be obtained (for an ensemble of 2^{nM} systems) [506]:

$$p \leqslant 2^{-n(\gamma_0-\gamma_n)} \tag{3.12}$$

where γ_0 is the estimate index defining the limiting transmission rate, γ_n is the transmission rate in bits per measurement (sampling)

$$\gamma_n = R/D \tag{3.13}$$

Here $D = n/T$, and $n = aFT$ is the number of independent samplings (measurements) in a channel with a bandwidth F. The coefficient a depends on the channel characteristics (in the limit, by the sampling theorem, $a = 2$ and $D = 2F$).

It follows from Eq. (3.12) that the average error probability can be made arbitrarily small by selecting sufficiently large n, if the specific transmission rate γ_n is less than the limiting rate γ_0. For the systems with binary signals [506] we have

$$\gamma_0 = 1 - \log_2 [1 + \exp(-E/N_0)] \tag{3.14}$$

where E is the energy of the signal, N_0 is the spectral density of noise.

Consequently, the most important parameter of a communication system is the specific transmission rate. The limiting value of the rate, γ_0, characterises a discrete channel (an analog channel with a modem), and γ_n characterises the possible rate interval that depends on the code selected. In actual calculation the specific rate is determined relative to the bandwidth of the channel [109] rather than to the number of measurements,

$$\gamma = R/F = a\gamma_n \tag{3.15}$$

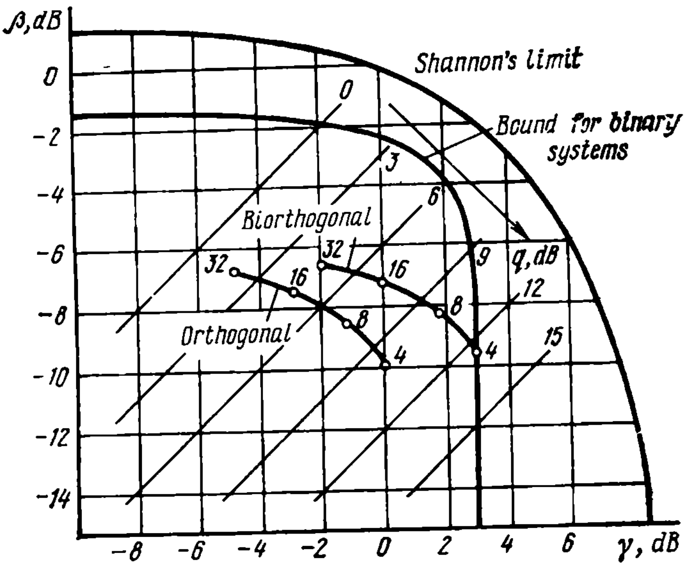

Fig. 3.2. Limiting efficiency curves

The quantity γ characterizes the utilization of the channel band (frequency efficiency). A second characteristic of the system is its power efficiency

$$\beta = RN_0/P_s = \gamma/q \tag{3.16}$$

where $q = P_s/P_n$ is the S/N ratio in the channel.

A general characteristic of the system is the channel capacity utilization (information efficiency)

$$\eta = R/C = \gamma_n/C_n = \eta_n \tag{3.17}$$

where, according to Shannon, [478],

$$C_n = C/D = 0.5 \log (1 + P_s/N_0F) \tag{3.18}$$

From Eqs. (3.17) and (3.18) we obtain

$$\eta = \gamma/\log (1 + \gamma/\beta) \tag{3.19}$$

Also, according to Shannon [478] the maximum of the function $\eta_{max} = 1$ is attained with appropriate methods of transmission (coding and modulation) and reception (demodulation and decoding). In doing so, the error can be made arbitrarily small, and we get

$$\beta = \gamma/ (2^\gamma - 1) \tag{3.20}$$

Equation (3.20) can be represented diagrammatically (Fig. 3.2). It is to be noted that the γ-effectiveness may vary within broad limits (from 0 to ∞ in analog transmission, and from 0 to $2\log m$ in

discrete transmission), whereas the β-effectiveness is limited from above, i.e. as $F \to \infty$, $\beta \to 1/\ln 2$. If represented on a logarithmic scale, the lines of equal S/N ratios are straight lines with a slope of 45°. The limiting expression for systems with binary signals (binary sequences) is obtained directly from Eq. (3.14)

$$\beta_n = \eta_n / \ln (2^{1-\gamma_n} - 1) \tag{3.21}$$

This variation is also given in Fig. 3.2.

In actual systems the error is always finite and $\eta < 1$. In these cases, at p = constant, the values of β and γ may be obtained individually and the curves $\beta = f(\gamma)$ may be plotted. These curves are also provided in Fig. 3.2. In the coordinates β vs. γ, to each form of an actual system there corresponds a point (or curve) in a plane. All these points (curves) are arranged below the Shannon's curve. The character of the curves depends on the type of signals (modulations), the code and signal handling technique. By way of example, Fig. 3.2 shows the limiting efficiency curves for systems with reception "as a whole" of orthogonal and biorthogonal signals. The numerals on the curves indicate the number of signal positions. The nomographs thus obtained enable one to identify with relative ease the systems which meet the specifications.

Analysis of effectiveness of certain systems with correcting codes. The most general way of improving the communication system efficiency both with simple (narrowband and wideband) signals, and with composite signals, is through increasing the number of signals used in transmission. But the possibilities of this method are substantially limited by the difficulties involved in implementing the hardware for the optimal processing of wideband signals, and by the impairment of the noise-immunity of narrowband signals.

These difficulties may be obviated using the correcting codes. The redundancy thus introduced is employed to detect and correct errors which occur due to interference as the preliminary decision is made on symbols (elements) of the signal. Unlike the optimal reception "as a whole", the elementwise reception allows the receiver to be drastically simplified, and with a large number of signals (code combinations) used in transmission it is in most cases the only possible reception. The punishment for such a simplification is energy loss of about 2 dB.

Figure 3.3 shows the efficiency curves for block coding systems with elementwise reception (algebraic decoding). It also presents for comparison an efficiency curve for systems with biorthogonal signals received "as a whole" (analog maximum a posteriori probability (MAP) decoding). The coefficients β and γ are for the most efficient block codes liable to be implemented in a relatively simple way,

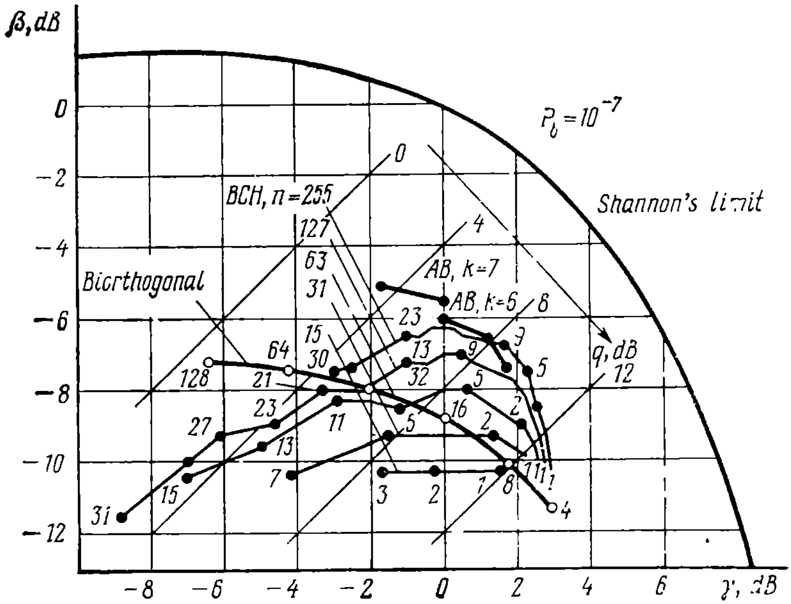

Fig. 3.3. Efficiency of systems with block codes

i.e. BCH codes (Kurtov, N. N., "On the efficiency of error-free coding in channels with multipositional signals", TUIC, Issue 84, 1977).

As is seen in the figure, higher β- and γ-efficiencies in the systems with correcting codes are achieved by increasing the block length. We know that here for most blocks, BCH included, the decoding complexity grows exponentially with increasing block length.

At present, codes are available which are simple to implement and whose decoding complexity grows slower than by exponent. These codes include, for instance, the low-density Gallager code. But the efficiency of these codes is low.

It is possible to avoid excessive complications of decoders as the code length increases by combining relatively short codes allowing a concatenated connection of decoders of initial codes. Figure 3.4 shows the efficiency curves for the systems with concatenated codes (dotted line) whose inner code here is the biorthogonal one with reception "as a whole", namely the M-ary RS code (M is the number of biorthogonal signals received "as a whole").

As the length of the RS code is limited by its basis, the efficiency of concatenated codes with small M is low. Any increase in M with the aim to use them in combination with long RS codes, apart from a drastic reduction in γ-efficiency, also results (at $M \geqslant 32$) in a significant complication of the modem. Under these conditions,

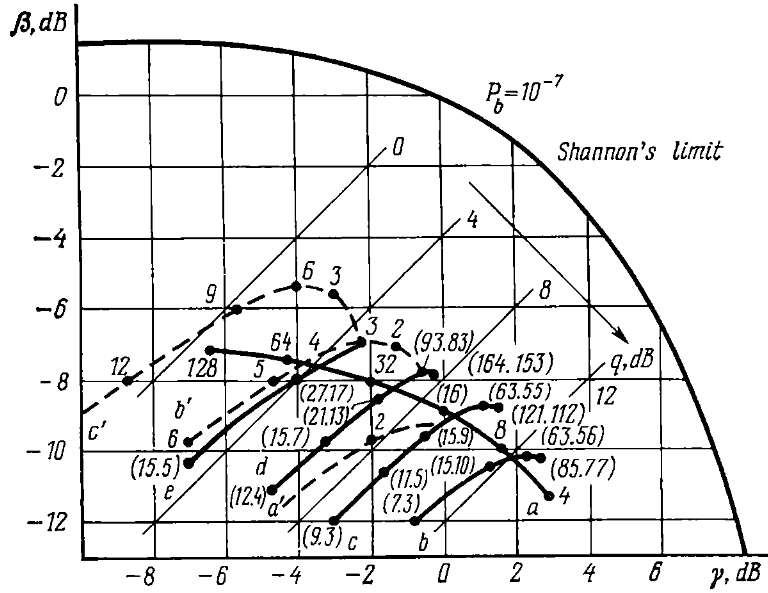

Fig. 3.4. Efficiency of systems with (a) biorthogonal signals, (a', b', c') concatenated codes, (b, c, d, e) outer binary codes correcting error packets

concatenated-type systems are of interest, in which a discrete widened channel formed by the initial continuous modem of the composite (biorthogonal) signals is taken to be a binary one with memory. A distinctive feature of such a channel is the known error statistics determined by the modulation system. This simplifies the task of matching the codes to the channel, the latter being one of the major causes of the limited application of codes in real systems.

The outer codes here are the binary block codes which correct error packets of known length defined by the modulation system, $b \leqslant \log_2 M$. The parameters of these codes (length is the first number, the number of information digits in a code combination is the second) are given in parenthesis along the curves. It is seen that these systems are more effective than those with concatenated codes at $M < 16$ and provide an added gain in β-efficiency as compared with the biorthogonal signals with γ-efficiency reduced but slightly. In this case, the outer binary code should have length $n > 60$.

There are also other ways to simplify the decoder realization, e.g. convolution coding. Worth especial mention are stochastic algorithms for decoding of these codes, and in the first place, the Viterbi algorithm. In contrast to many others, this algorithm not only realizes the optimal decoding using the MAP decision criterion, but employs

a flexible decision in the process, i.e. the quantization of channel symbol correlator for a number of levels more than two. This allows the code parameters to be matched to the modulation system in order to improve the efficiency of the communication system as a whole. Figure 3.3 depicts the efficiency curves AB for the systems using the convolution codes with the constraint length $\nu = 6$ and $\nu = 7$ decoded by the Viterbi algorithm with flexible decision. As is seen in the figure, the efficiency of such systems is higher than that of block coding systems of similar complexity.

Chapter

4

Pseudonoise Signals

4.1. General

Noise immunity of communication systems with PNSs. The pseudonoise, or complex, signals (PNS) are such signals for which the product of bandwidth F by duration T is much larger than unity. The product

$$B = FT \tag{4.1}$$

we shall call the base of PNS. The base characterizes both the complexity of the signal proper and the complexity of the radio communication system using PNSs. Use of PNSs in communication systems allows for improving the noise immunity and secrecy of information transmission, and to combat the path multiplicity. PNSs are employed to separate users in asynchronous address communication systems (AACS).

The signal-to-noise ratio at the receiver output, $h^2 = (P_s/P_{int})_{out}$, which characterizes noise immunity for PNS systems, is given by

$$h^2 = (P_s/P_{int})_{in} B \tag{4.2}$$

with the signal energy

$$E = P_sT \tag{4.3}$$

and the power spectral density of noise

$$N_{int} = P_{int}/F \tag{4.4}$$

is constant within a frequency band occupied by the signal spectrum. Figure 4.1 shows the noise immunity curves for a PNS communication system in coherent detection of two opposite signals with a base $B = 100$, for an FM communication system with equivalent base 100, and for an AM communication system with equivalent base 1. As is seen in the figure, NLS communication systems are most advisable for $(P_s/P_{int})_{in} \ll 1$, i.e. under conditions when traditional

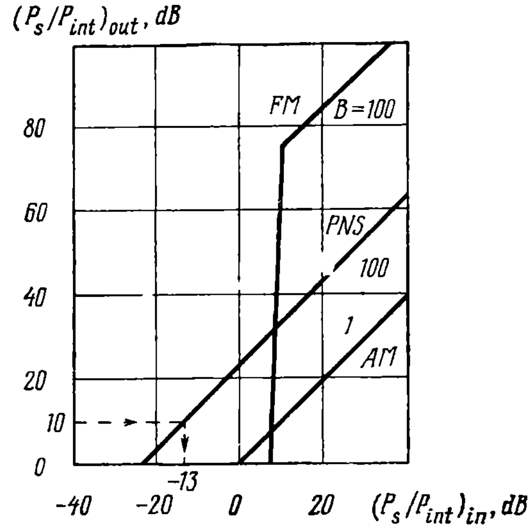

Fig. 4.1. Noise immunity of PNS, FM and AM communication systems

systems fail to provide reliable transmission of information. In present-day systems the base may be from 10^3 to 10^6.

Bandwidth and capacity. With a given information transmission rate, larger bases are only possible at the expense of a wider signal spectrum. Suppose that for the transmission of a discrete information an alphabet of m signals is used, each of which being a PNS with base B. Given the transmission rate R and error probability per one binary digit, P_{e2}, then with incoherent reception of m signals the required bandwidth is given by

$$F_m = R\,(P_{int}/P_s)_{in}\,[1 - \ln(4P_{e2})/\ln m]\,2\ln 2 \tag{4.5}$$

For larger alphabet m, the required bandwidth is reduced and in the limit as $m \to \infty$ from Eq. (4.5) we obtain

$$R = (F_m/2\ln 2)\,(P_s/P_{int})_{in} \tag{4.6}$$

The number of orthogonal signals is known to be equal to the base, i.e. $m = B_m$, where

$$B_m = (F_m/R)\log_2 m = (F_m/R)\log_2 B_m \tag{4.7}$$

It follows from Eqs. (4.6) and (4.7) that for a specified $(P_s/P_{int})_{in}$ the necessary base value of an mth PNS is related to the alphabet by the equation

$$B_m/\ln B_m = 2\,(P_{int}/P_s)_{in} \tag{4.8}$$

Noise immunity with heavy structural interference. The action of such interference is decided by the cross-correlation function (CCF)

of the information-bearing signal and the structural interference that belongs to the same class as the information-bearing signal. It is normally assumed that

$$|R_{\text{CCF}}| = \sqrt{\alpha/B} \tag{4.9}$$

where α is a constant pertaining to the increase in CCF with reference to the level $1/\sqrt{B}$. The maximum peak of CCF is

$$R_{\text{CCF, max}} = \sqrt{\alpha_{\max}/B} \tag{4.10}$$

If the CCF maximum coincides with the instant of decision making, then the signal-to-structural-interference ratio will be

$$\begin{aligned} h^2 &= (P_s/P_{int})_{in}\, B\,(1 - R_{\max})^2 \\ &= (P_s/P_{int})_{in}\, B\,(1 - \sqrt{\alpha_{\max}/B})^2 \end{aligned} \tag{4.11}$$

Equation (4.11) is similar in structure to Eq. (4.2). By Eq. (4.11), the availability of the quantity $\alpha_{\max}$ is responsible for a decrease in the signal-to-structural-interference ratio at the output. The signals used in the systems should have small peaks of CCF with all possible signals constituting the complete code. And it is one of the conditions that are to be met in designing PNSs. To evaluate the influence of CCF peaks for any arbitrary pair of signals, we assume that they are normally distributed with the variance $1/B$. The probability that the CCF modulus exceeds a certain level R_{CCF0} is

$$\mathrm{P}_{ex} = 1 - [1 - \mathrm{p}\,(R_{\text{CCF0}})]^B \tag{4.12}$$

where the probability for one CCF peak to exceed the level will be given by

$$\mathrm{p}\,(R_{\text{CCF0}}) = 2\,[1 - F\,(R_{\text{CCF0}}\sqrt{B})] \tag{4.13}$$

and $F(\cdot)$ is the normal distribution function. The dependence on R_{CCF0} in Eq. (4.12) is of a threshold nature. For $R_{\text{CCF0}} > R_{thr}$, the probability P_{ex} tends to zero, and at $R_{\text{CCF0}} < R_{thr}$, P_{ex} approaches unity. The threshold value is

$$R_{thr} = \sqrt{2 \ln B/B} \tag{4.14}$$

Noise immunity with combined interference. In recent years, great attention has been paid to jamming suppression in communication systems, with the objective to provide continuous performance of the system under conditions of combined interference. For example, the following disturbances may occur simultaneously: noise interference (intrinsic noise of the receiver), jamming noise, pulsed jamming, structural interference of the system, etc. The problem of combined interference was discussed by many investigators. They mostly treat effects of noise and narrowband interference [9, 242, 243, 265-267, 299].

The combined use of rejector (band-pass) filters and wideband clipping amplifier—matched filter systems forms the basis for a number of heuristic methods for suppressing the narrowband pulsed noise, and structural interference. Along with the heuristic methods of synthesis of receivers invariant to combined interference, there are conventional techniques of synthesis of adaptive receivers (see, e.g., [163, 164, 202, 226, 288, 296, 355]).

4.2. PNS Communication Systems

Synchronous address communication systems. In multichannel synchronous communication systems the efficiency of use of the channel bandwidth is drastically dependent on the activity of users. If the latter is sluggish, then, in addition to classical time and frequency division multiplex (TDM and FDM) of user messages, the code division technique (CD) [70, 271] is adopted. With CD methods, discrete information is transmitted in the form of address code combinations which here are PNSs, such as PSK signals that are orthogonal or quasiorthogonal. The employment of PNSs in synchronous addressing communication systems (SACS) has the advantage that it improves the use of band and time allocated for a channel, provides communication for more users at low activity. But when CD is used with an arbitrary number of overlapping address signals, a group signal is a random process in time, and therefore the operation of a transmitter (by power) may be less efficient. To improve efficiency of an PNS-SACS, recourse must be made either to AGC of a group signal, or to rigid limitation of a group signal [288]. Studies indicate that the strict limitation technique is inferior to the adaptive method with AGC of a group signal as far as the noise immunity is concerned.

Denote by L_{ch} the number of channels in linear TDM or FDM whereby each user is allocated its own time slot or frequency channel, respectively. For instance, with frequency division $L_{ch} = F/F_u$, where F is the total bandwidth of the channel, F_u is the bandwidth of a user's channel, and $F_u = \beta F_{com}$, where F_{com} is the width of the communication channel, coefficient β taking into account the width of guard bands, $\beta > 1$. Let h_L^2 be the signal-to-noise ratio at the output of one channel in linear separating, and $h_0^2 = P_s T/N_0$ be the signal-to-noise ratio at the output of a group circuit for a maximum power of the receiver. In that case

$$h_L^2 = h_0^2/L_{ch} \tag{4.15}$$

When AGC with l_u active users, h_{AU}, is used

$$h_{AU}^2 = h_0^2/l_u = h_L^2 L_{ch}/l_u \tag{4.16}$$

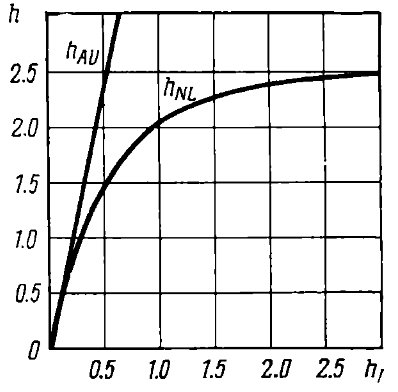

Fig. 4.2. Noise immunity of PNS-SACS

and the mean number of active users is $\bar{l}_u = p_u L$, where p_u is the probability that the channel is occupied by the users or the activity of the users, L is the total number of users in a PNS-SACS. If $\bar{l}_u \approx L_{ch}$, then the SACS with AGC provides the S/N ratio as the SACS with the linear multiplexing, but admits to the communication channel by far a larger number of users. At $L = L_{ch}$, the AGC-SACS shows a better noise immunity.

In a SACS with a strict limitation of the group signal (non-linear SACS) the ratio of signal to noise plus interference will be

$$h_{NL}^2 = [2N/\pi (l_u - 1)] [h_0^2/(2h_0^2 + N)] \tag{4.17}$$

where N is the number of pulses in a PSK signal. For a small S/N ratio, $N \gg 2h_0^2$ and

$$h_{NL}^2 \cong \frac{2h_0^2}{\pi (l_u - 1)} \approx \frac{2}{\pi} h_{AU}^2 = \frac{2}{\pi} \frac{L_{ch}}{l_u} h_L^2 \tag{4.18}$$

i.e. the non-linear SACS is $2/\pi$ times worse than the adaptive system, but $2L_{ch}/\pi l_u$ better than the linear system. Figure 4.2 depicts the dependences of h_{AU} and h_{NL} on h_L for $p_U = 0.05$. As is seen in the figure, the improvement is only possible for small S/N ratios. But in that case the linear SACS may be rendered nonreliable. The adoption of rigid limitation gives no improvement, as the modulus of the signal-to-noise ratio, h_{NL}, is just larger than unity. It follows from the figure that the adaptive SACS compares favorably with the others.

Asynchronous address communication systems. Asynchronous address communication systems (AACS) rely on the code division of users, whereby each user is allocated its own (user's) alphabet of signals (or code sequences) to transmit information. Here the division is possible because signals from various users differ in waveform.

With this method of separation, the information being transmitted is supplied with an "address" represented here by specified signals. It is the availability of "addresses" that enables the asynchronous regime of simultaneous functioning of many users to be achieved.

The pioneering work in the code multiplexing and division has been that by Ageev [2]. This work introduces the foundations of linear separation by using linearly independent signals. In linear separation there are no structural (interchannel) interferences. In code division, in AACS there occur structural interferences as a result of simultaneous operation of users within a common bandwidth. With code division, however, the parameters of signals can be chosen so that the level of structural interference will be arbitrarily small. i.e. a preselected noise immunity will be provided.

To provide an asynchronous communication, it was sufficient to use the frequency division, whereby every user was allocated his own channel. It appeared later that the frequency division was not able to guarantee asynchronous radio communication with a moving object. It is for these reasons that developments of AACS [454] were initiated. The AACS was brought to the forefront by the work of Kostas [345] which highlighted the basic features of such systems, including the electromagnetic compatibility of wideband and narrowband systems of information transmission (ITS). Then a number of publications followed in which both theoretical and engineering design aspects of AACS were considered [40, 41, 49, 50, 53, 66, 70, 84, 210, 220, 222, 224, 326]. Reference [70], which was the first book devoted to AACS, should be marked out especially. At that time a large number of various AACSs designed for both terrestrial and satellite communication systems were proposed. That period is to be viewed as the earliest period of AACS development. Their advantages were evident, whereas the disadvantages seemed insignificant. Over the years the disadvantages of the AACSs became more conspiquous with the result that the AACSs began to attract less attention, and the publication rate on the subject dwindled accordingly. Now, the second period of AACs development seems to have started (see, e.g. [72, 180, 227, 249, 289-292, 301-304, 497, 498]). Particular significance is attached today to code division in the earth's and satellite AACSs to attack the problems in air traffic control (ATC) and electromagnetic compatibility.

With code division, the number of signals required for an AACS is equal to the product of the number of users by the number of signals in the alphabet (we assume that all the users use alphabets of similar size). The minimum number of signals equals the number of users. Should the users be numerous, the selection of the signals will be one of the key issues in the development of an AACS.

Denote the number of users in the AACS by L_{ch} and the number of active users transmitting information at a given instant of time

by l_u. The input of the receiver of one of the users receives signals coming from $l = l_u - 1$ disturbing users which cause crosstalk, and an information-bearing signal from the active user that transmits information to the given user. If $l \gg 1$, then $l \approx l_u$.

Suppose now that the bandwidth for all the signals in the AACS is equal to the total bandwidth F. We further assume that all the active users produce at the input of the selected jth receiver signals of equal power P_s. In that case, the signal-to-structural interference ratio will be given by

$$h^2 = P_sT/N_{int} = P_sFT/lP_s = B/l \tag{4.19}$$

Equation (4.19) suggests that for a specified l, improved S/N ratio might be only achieved by using a larger base of signals.

In the presence of structural interference and noise with spectral density N_0, the signal-to-(structural interference plus noise) ratio [84] will be

$$h^2 = (l/B + 1/h_0^2)^{-1} \tag{4.20}$$

or after rearrangement

$$lh^2/B = 1 - h^2/h_0^2 \tag{4.21}$$

where $h_0^2 = P_sT/N_0$.

Correlation technique gives accordingly [50]

$$h^2 = (2\alpha l/B + 1/h_0^2)^{-1} \tag{4.22}$$

In the derivation of Eq. (4.22) it has been assumed that the root-mean-square value of CCF has the form

$$\sigma_{R\ \mathrm{CCF}} = \sqrt{\alpha/B} \tag{4.23}$$

Equations (4.20) and (4.22) yield the same result at $\alpha = 1/2$. For real signals, $\alpha > 1/2$.

With simultaneous action of jamming and structural interference, we obtain

$$h^2 = (P_s/P_{int})_{in} B/[1 + (P_s/P_{int})_{in} 2\alpha l] \tag{4.24}$$

In some communication systems, discrete information is transmitted asynchronously. Here an overlap of k PNSs might be supposed. The signal-to-(noise plus own structural interference) ratio is

$$h^2 = (P_s/P_{int})_{in} B [1 + (P_s/P_{int})_{in} 2\alpha (k - 1)]^{-1} \tag{4.25}$$

The utilization efficiency of the total bandwidth is characterized by the capacity factor

$$\mu = l_u F_s/F \tag{4.26}$$

and by the number of active users per unit band

$$\gamma = l_{u\ \max}/F \tag{4.27}$$

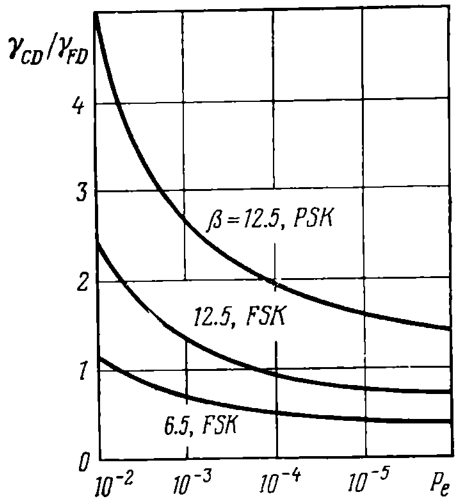

Fig. 4.3. Efficiency of AACS

where $l_{u\ max}$ is the maximum number of active users operating in the total band with a specified quality. For CD-AACS we have

$$\mu_{CD}/\mu_{FD} = \beta/p_u h^2; \quad \gamma_{CD}/\gamma_{FD} = \beta/h^2 \tag{4.28}$$

Figure 4.3 presents the dependences of γ_{CD}/γ_{FD} on P_e for various $\beta = F_u/F_s$ and reception methods. If $h^2 < \beta$, then CD compares favorably with FD. It is shown in reference [55] that the adoption of CD in rural radio communication would allow solving the problem of providing radio communication for a great many users without the need to allocate additional bands.

If structural interference and noise occur simultaneously

$$\gamma_{CS} = (1/2\alpha)\,(1/h^2 - 1/h_0^2) \tag{4.29}$$

where h^2 is obtained from Eq. (4.22), and $h_0^2 = P_s T/N_0$ is the S/N ratio. The efficiency of AACS in transmitting continuous and discrete messages, using various reception techniques, is dealt with in detail in [220-222].

4.3. Sets of Pseudonoise Signals

Classification. In large-base PNS communication systems, three classes of signals have found their application:

1. Phase-shift keying (PSK) signals (signals with code phase modulation—CPM signals).
2. Discrete frequency (DF) signals (signals with code frequency modulation—CFM signals).

3. Multiple discrete frequency (MDF) signals (multiple signals with code frequency modulation—MCFM signals).

Each time the second variant of discrete signal classification is shown in brackets.

Phase-shift keying signals are a sequence of radio pulses with phases varying according to a predetermined law. Normally the phase takes on the values (0 or π). In that case a PSK video signal composed of positive and negative pulses corresponds to a radio-frequency PSK signal. If the number of pulses is N, then the duration of one pulse will be $T_0 = T/N$, and the width of its spectrum is approximated by the signal spectrum $F = 1/T_0 = N/T$.

Discrete frequency signals are simply a sequence of radio pulses with carriers varying by a specified law. Let the number of pulses in a DF signal be M, pulse duration $T_0 = T/M$, its spectral width $F_0 = 1/T_0 = M/T$, and the base of DF signals $B = M^2$.

Multiple discrete frequency signals are DF signals in which every pulse is substituted by a pseudonoise signal.

Correlation properties and volumes. Correlation properties of any two signals with equal energies are decided by their CCF.

$$R_{\text{CCF}\ jk}(\tau) = \frac{1}{2E} \int_{-\infty}^{\infty} U_j(t)\, U_k^*(t-\tau)\, dt \tag{4.30}$$

Here $U_j(t)$ and $U_k(t)$ are complex envelopes of signals. For PSK signals, CCF takes the form

$$R_{\text{CCF}\ jk}(m) = \frac{1}{N} \sum_{n=m+1}^{N} a_{jn} a_{k,\, n-m} \tag{4.31}$$

where a_{jn}, a_{kn} are the symbols of two code sequences for signals with subscripts j and k. Eq. (4.31) described the case of aperiodic conditions, where the sequences overlap in part. For periodic conditions, CCF will be given by

$$R_{\text{CCF}\ jk} = \frac{1}{N} \sum_{n=1}^{N} a_{jn} a_{kn} \tag{4.32}$$

Under periodic conditions, two sequences may be displaced about each other, but Eq. (4.32) always has N terms.

According to the volume L, the signal systems can be divided into three classes: small, when

$$L = \sqrt{B} \ll B \tag{4.33}$$

normal (orthogonal or quasiorthogonal), when

$$L = B \tag{4.34}$$

and large, when

$$L \gg B \tag{4.35}$$

Most of known systems of signals are small or normal. For present-day communication systems, PNS sets are needed with the potency

$$L = Ae^{\gamma B} \tag{4.36}$$

where A, γ are constants. This law being difficult to realize, large systems should be developed with the potency increasing by the power law

$$L = AB^n \quad \text{at } n \gg 1 \tag{4.37}$$

Limits are known for any large PNS—so-called complete codes. The complete code is a system of signals, including all the signals of a class for the specified alphabet and the number of symbols in a signal. The alphabet of symbols is a set of various symbols comprising a signal. The complete code cannot be expanded as it exhausts all possible signals. As any PNS set is a subset of its complete code, it should possess some properties of a complete code. The larger the system, the closer its approximation to the complete code. The complete code of PM signals contains

$$L = p^N \tag{4.38}$$

code sequences, where N is the length of a code sequence, p is the number of symbols in the alphabet.

The studies of various properties of the complete binary code ($p = 2$) are given in references [17, 60, 85, 94, 188, 189, 223, 260]. It has been demonstrated in reference [56] that the variance of autocorrelation function (ACF) of the complete binary code is

$$\sigma^2 = 1/2N \tag{4.39}$$

and the coefficient of excess

$$\gamma_{exc} = 1 - 4/N + 2/N^2 \tag{4.40}$$

at $N \gg 1$; $\gamma_{exc} \approx 1$.

Correlation properties of PSK signals. In the analysis of correlation properties of PSK signals, the model of random processes is frequently applied, assuming that the symbol a_{jn} with equal probabilities takes on the values $+1$ or -1 (see, e.g. [46]). In that case the values of R of CCF or ACF are random values. This way of investigating the correlation behaviour of PSK signals is equivalent to the consideration of properties of a complete code.

The results of references [46, 94, 188] indicate that the distribution of periodic CCF of PSK signals should be described by a bino-

mial law

$$\mathrm{P}(R_{\mathrm{CCF}}) = C_N^{0.5(1+R_{\mathrm{CCF}})N} 2^{-N} \tag{4.41}$$

which at a large N is approximated by a normal law

$$w(R_{\mathrm{CCF}}) = \sqrt{N/2\pi} \exp(-NR_{\mathrm{CCF}}^2/2)$$
$$\text{at} \quad -1 \leqslant R_{\mathrm{CCF}} \leqslant 1 \tag{4.42}$$

The quantity $|R_{\mathrm{CCF}}|$ is distributed according to the law

$$w(|R_{\mathrm{CCF}}|) = \sqrt{2N/\pi} \exp(-NR_{\mathrm{CCF}}^2/2)$$
$$\text{at} \quad 0 \leqslant R_{\mathrm{CCF}} \leqslant 1 \tag{4.43}$$

The distribution of aperiodic correlation functions for PSK signals might approximately be assumed to be normal

$$w(R_{\mathrm{CCF}}) = \sqrt{N/\pi} \exp(-NR_{\mathrm{CCF}}^2)$$
$$\text{at} \quad -1 \leqslant R_{\mathrm{CCF}} \leqslant 1 \tag{4.44}$$

In the first approximation, the value of $|R_{\mathrm{CCF}}|$ under aperiodic conditions is distributed by

$$w(|R_{\mathrm{CCF}}|) = 2\sqrt{N/\pi} \exp(-NR_{\mathrm{CCF}}^2)$$
$$\text{at} \quad 0 \leqslant P_{\mathrm{CCF}} \leqslant 1 \tag{4.45}$$

In reference [482] the lower bound of the volume of PSK signal system is determined. This bound was derived with the use of the Chernov bound to approximate the distribution of correlation functions by a binomial distribution. But the Chernov inequality provides a good approximation for the "wings" of probability density functions only. This means that the volume derived with the Chernov bound relates to the signal systems with "poor" correlation behavior. Therefore it is necessary to use the approximation by the power law that gives sufficient accuracy with relatively small values of R_{CCF}. In approximation of the CCF probability density by the Edgeworth series with an excess coefficient $\gamma = 1$, the mean power $\overline{L}$ of a PSK signal set satisfies the inequality

$$\overline{L} \geqslant (3\sqrt{\pi}/4R_{\mathrm{CCF0}}^3 N^{5/2}) \exp(R_{\mathrm{CCF0}}^2 N) \tag{4.46}$$

where R_{CCF0} is the permissible maximum value of the CF modulus. The signal set with power $\overline{L}$ is comprised by a system with power $L > \overline{L}$, from which the signals possessing CCFs and ACFs with $R_{\mathrm{CCF}} > R_{\mathrm{CCF0}}$ are removed. The threshold level will be

$$R_{thr} = \sqrt{\ln(aN)/N} \tag{4.47}$$

where $a = 2^{3/2}\pi^{-1/2}$. At $R_{CCF} < R_{th}$ the probability that R_{CCF0} be exceeded is practically 1, and at $R_{CCF} > R_{th}$ the probability that the R_{CCF} peak exceeds R_{CCF0} drops drastically. Therefore we shall express R_{CCF0} in relative units as related to R_{th}. We assume that

$$R_{CCF0} = \sqrt{\nu \ln (aN)/N} \tag{4.48}$$

Then

$$\overline{L} > c(\nu)\,[\ln (aN)]^{-3/2} N^{\nu-1} \tag{4.49}$$

where $c(\nu) = 3\pi^{1/2} a^{-\nu} 2^{-2} \nu^{3/2}$.

It is shown in reference [52] that for optimal PSK signals the number of blocks μ is equal (exactly or approximately) to

$$\mu_0 = (N + 1)/2 \tag{4.50}$$

where N is the number of pulses in a PSK signal, and the block is a sequence of identical pulses. Such a property warranted a hypothesis: optimal PSK signals are to be found among a set of PSK signals complying with the equality $\mu \approx \mu_0$, i.e. this equality is a nécessary condition for PSK signals to be optimal (see also [59, 61]).

Multiplicative and segment signal sets. Most of known PSK signal systems are normal (i.e. for them $L = \beta$) and either multiplicative or segmentary. The multiplicative systems consist of multiplicative PSK signals obtained by symbol-by-symbol multiplication of two PSK signals. Of special value among them are the systems constructed as follows. For the initial system, a certain system of signals is used, whose correlation behavior is not good, but which shows promise from the standpoint of simplicity of formation and processing. Then a producing signal possessing adequate correlation properties is chosen. By multiplying the generating signal by every signal of the initial system, we arrive at a multiplicative system. In [237] the Walsh system (the Read-Müller code) is employed, every signal of which is multiplied by the same predetermined generating signal. As a result a new system of signals is produced, in which the side lobe CCF peaks are on the average lower than in the initial system.

Segment systems are those formed from the segments of M-sequences. The segment system is a derivative one, as the segregation of a segment from an M-sequence is equivalent to the use of a narrow-band producing signal, i.e. a simple signal with a square envelope, the duration of which being equal to that of the segment (see also [143]). In [57] the technique was put forward for determining the length of a segment and the volume of a signal system, proceeding from specified correlation properties. References [328, 431] are also devoted to studies of properties of segment signals.

It is to be noted that segment and multiplicative signals comply with equality (4.50). The same property exhibit also the maximum

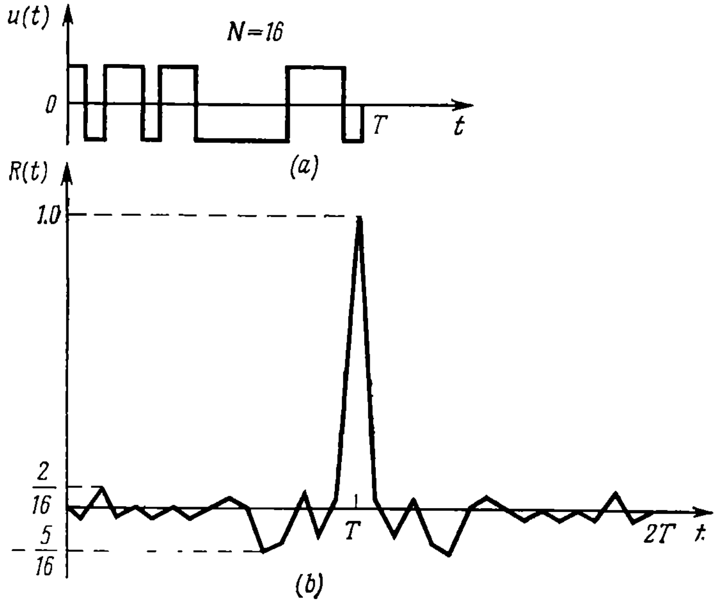

Fig. 4.4. Maximum probability sequence (*a*), and its autocorrelation function (*b*)

probability sequences (MPS) proposed in [62]. These sequences are formed by a random or determinated method from blocks for which the length k and number Λ_k are chosen from the condition

$$\delta = \sum_{k=1}^{\log_2 \mu_0} k\,|\Lambda_k - \overline{\Lambda}_k| = \min_{\Lambda_k} \qquad (4.51)$$

where

$$\overline{\Lambda}_k = \mu_0/2^k \qquad (4.52)$$

The number of blocks with length k is proportional to the probability, equal to $1/2^k$, of occurrence of such a block in a random signal. Fig. 4.4 shows an MPS with $N = 16$ and $\mu_0 = 8$, and corres-

TABLE 4.1

Sequence type	Statistics		
	$\lvert R_{CCF\,max} \rvert \sqrt{N}$	$m_1(\lvert R_{CCF} \rvert)\sqrt{N}$	$\sigma(\lvert R_{CCF} \rvert)\sqrt{N}$
MPS	0.75-2.0	0.35	0.33
M-sequence	0.7-1.25	0.32	0.26
Segment sequence	1.45-4.1	0.52	0.9
Random sequence	1.5-3.1	0.51	0.65

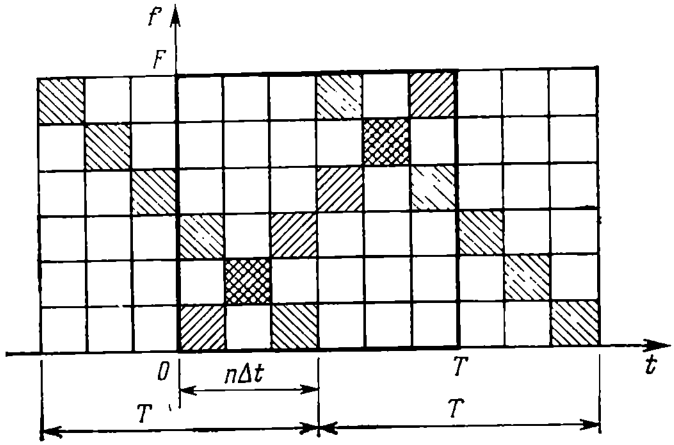

Fig. 4.5. Coincidence of two discrete frequency signals in frequency-time plane

ponding ACF. Table 4.1 gives characteristics of MPS, M-sequences, segment and random sequences.

It follows from the table that MPSs show statistical behavior close to that of the best sequences, *viz.* M-sequences, but the number of MPSs is markedly greater than the number of M-sequences. It is

$$L = (\log_2 \mu_0 + 1)! \prod_{m=1}^{\log_2 \mu_0 - 1} [(\log_2 \mu_0 - m)!]^{2m-1} \tag{4.53}$$

and grows abruptly with increasing μ. A regular technique of construction of large PSK signal systems is suggested in [193].

Discrete frequency signals. A characteristic feature of correlation functions of DF signals is their being completely defined by the number of coincidences of m elements in the frequency-time plane. In Fig. 4.5, one DF signal is shown by heavy lines and occupies the time interval $(0, T)$. It contains M elementary radio pulses. A second signal, i.e. a pair of disturbing signals, is shifted relative to the first one by the time Δt, which, for simplicity, is put to be a multiple of the element duration T_0. The signals are distinguished by the direction of shading. The two coincidences of DF signals correspond in Fig. 4.5 to the elements with double shading. The coinciding elements make to CCF of two DF signals a contribution in the form of the ACF of a square radio pulse, which has a triangular envelope with a maximum equal to $1/M$ and is located at a point with the time that is a multiple of T_0. If a simultaneous coincidence of m elements occurs, then the modulus of CCF for two DF signals at

points $\tau = nT_0$ (n is an integer) is defined by the inequality

$$|R_{\text{CCF}}| \leqslant m/M \tag{4.54}$$

Reference [58] gives regular algorithms for the construction of optimal DF signal systems. So, symbols $\alpha_j(\nu)$ of jth code sequence may be determined by the rule

$$a_j(\nu) \equiv C_0 a^{j+\nu} \pmod{M+1} \tag{4.55}$$

where $C_0 = \overline{1, M}; j, \nu = \overline{0, M-1}$, a is a primitive root. The volume of optimal systems of DF signals is $M = \sqrt{B}$, i.e. such systems are small.

In work [64] the distribution of the number of coincidences for aperiodic CCF is discussed, and the probability of m coincidences, at the discrete shift equal to n, is

$$\mathrm{P}_{M,n}(m) = n!\, D_{M,m,n}/M! \tag{4.56}$$

where

$$D_{M,m,n} = C_{M-n}^{m} D_{M-n,0,n} \tag{4.57}$$

$$D_{M,0,n} = (M!/n!)\,[1 - C_{m-n}^{1}/M + C_{m-n}^{2}/M(M-1) - \ldots + (-1)^{M-n}] \tag{4.58}$$

From Eq. (4.56), we have at $n < 0.6\ M$

$$\mathrm{P}_{M,n}(m) \approx \frac{C_{M-n}^{m}}{C_{M}^{m}} \frac{1}{em!} = \frac{C_{M-n}^{m}}{C_{M}^{m}} \mathrm{P}_M(m) \tag{4.59}$$

where $\mathrm{P}_M(m)$ is determined from Eq. (4.56) for $n = 0$. It follows from Eq. (4.59) that $\mathrm{P}_{M,n}(m)$ is close to $\mathrm{P}_M(m)$ for small n.

Using the method put forward in [482] it might be shown that the lower volume boundary of large DF signal systems complies with the following inequality:

$$\overline{L} > 0.25\,(n+1)!\, MM!\, [MM!\, e^{-1} + (n+1)!\, e^{M}]^{-1} \tag{4.60}$$

where n is the maximum number of coincidences. For $n \ll M$, we obtain

$$\overline{L} > 0.25\, e\, (n+1)! \tag{4.61}$$

The discrete composite signals (DCF-PSK, DCF-FM, DCP-PSK) have been discussed by many authors [54, 65, 83, 204, 205, 206, 232, 248, 264]. They have the distinctive feature that their energy is concentrated in separate areas of the frequency-time plane, which makes it possible to use efficiently the multichannel frequency and time processing.

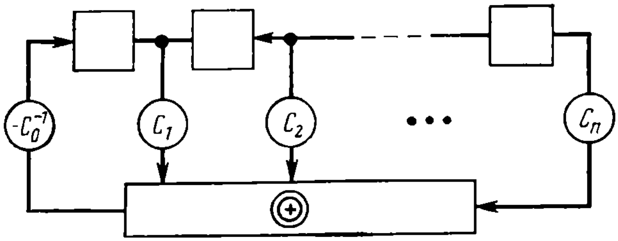

Fig. 4.6. Canonic configuration of linear autonomous automaton

4.4. Generation of PNSs with Predetermined Distribution of Symbols

Generation of M-sequences. Consider now the shaping of PNSs with a predetermined symbol distribution, using relatively simple transformations of binary and multisymbol M-sequences.

For a discrete (by amplitude and time) ergodic signal with independent values of

$$x(t) = x_i(\Delta t_i) \text{ for } x_i \in \{A_0, A_1, \ldots, A_{q-1}\} \tag{4.62}$$

the joint probability equals the product of probabilities of elements

$$p(x_i = A_{\nu, i}, x_{i+1} = A_{\nu, i+1}, \ldots,$$
$$x_{i+k-1} = A_{\nu, k-1}) = \prod_{j=1}^{k-1} p(A_{\nu, j}) \tag{4.63}$$

Then the ACF of the signal will be given by

$$R_{\text{ACF}}(\tau) = \begin{cases} 1 - \tau/t, & |\tau| \leqslant \Delta t \\ \alpha, & |\tau| > \Delta t \end{cases} \tag{4.64},$$

$$\text{where } \alpha = \frac{\left(\sum_j p(A_j) A_j\right)^2}{\sum_i p(A_j) A_j^2} \tag{4.65}$$

The task is to seek PNSs with correlation properties as close to the above ones as possible.

The theory of linear recurrent sequences has gained wide recognition [510]. Therefore, we here only discuss some principal properties important for our reasoning below.

Figure 4.6 presents a canonical configuration of a linear automaton. By a linear differential equation, every n elements of the sequence always combine to form the $(n + 1)$th element

$$a_i = -C_0^{-1} \sum_{j=1}^{n} C_j a_{i-j}, \; a_i, \; C_j \in GF(q) \tag{4.66}$$

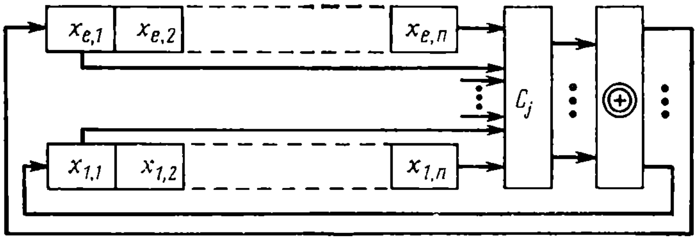

Fig. 4.7. Production of multiple-valued maximum sequences in binary-coded form

where $GF\ (q)$ is the Galois field, q is a prime number to the mth power, $q = p^m$.

It is a common knowledge that the necessary and sufficient condition for the linear automaton to produce a sequence with a maximum period

$$N = q^n - 1 \tag{4.67}$$

is that the characteristic polynomial

$$C(x) = C_0 + C_1 x + \ldots + C_n x^n,\ C_0,\ C_n \neq 0 \tag{4.68}$$

is prime and irreducible to $GF\ (q)$.

Physically, the linear automation for $p = 2$ is a flip-flop shift register with feedback through a modulo-2 summator. The electronic generation of multisymbol M-sequences makes it possible to simulate random signals and to solve other problems [342, 352, 363, 449].

Multisymbol M-sequences can be readily produced in binary coding, using integral circuits. The feasibility of generating M-sequences for $q = 3$, 5, and 7 and periods up to $N = 10^6$ has been amply confirmed experimentally [364]. The time delay of binary-coded elements of sequence is achieved owing to two or three binary shift registers. Coefficients C_j are formed and added up in modulo p using combinatorial logic circuits (Fig. 4.7).

Within one period of M-sequences, the repetition rate of any combination of k elements is given by [510]

$$\left.\begin{aligned} h\,(a_1 \ldots a_k) &= q^{n-k} \\ h\,(00 \ldots 0) &= q^{n-k} - 1 \end{aligned}\right\} \quad \text{at} \quad k \leqslant n \tag{4.69}$$

The relative joint repetition rate of M-sequences then corresponds to the joint probability of independent signals of individual symbols $p\ (A_\nu) = 1/q$

$$h\,(A_{\nu,1} \ldots A_{\nu,k})/(q^n - 1) \approx 1/q^k \tag{4.70}$$

The autocorrelation function of the sequence shows a striking similarity to ACF of independent random signals [see Eq. (4.64)].

Generation of PNSs with specified symbol distribution, using transformations of M-sequences. The mapping

$$(a_i,\ a_{i+d_1},\ \ldots,\ a_{i+d_{k-1}}) \to b_i \tag{4.71}$$

of appropriate k elements of q-nary M-sequence $\{a_i\}$ on the sequence element $\{b_i\}$ makes it possible to obtain a periodic sequence $\{b_i\}$, $b_i \in \{B_0,\ B_1,\ \ldots,\ B_{q_T-1}\}$ with relative repetition rates of symbols

$$\begin{gathered} h\,(B_0)=(l_0q^{n-k}-1)/(q^n-1)\approx l_0/q^k \quad \text{at} \quad k<n \\ h\,(B_j)=(l_jq^{n-k})/(q^n-1)\approx l_j/q^k \\ \text{at} \quad j=1,\ \ldots,\ q_T-1 \end{gathered} \tag{4.72}$$

The mapping of combinations of k elements of M-sequence onto a transformed sequence should meet uniqueness conditions.

It might be indicated that the sequence $\{b_i\}$ will approach a random independent sequence if the k elements of the combinations subject to mapping are characterized by a sufficiently large quantity [364]

$$(a_i,\ a_{i+d},\ \ldots,\ a_{i+(k-1)d}) \to b_i \quad \text{at} \quad dk \leqslant n \tag{4.73}$$

For $k = 1$, PNSs are obtained which have independent repetition rate up to the order of n.

Example

$$\begin{aligned} &\{a_i\} = \ldots,\ 44340233130411210322420 1,\ \ldots \\ &k=1,\ (0), \\ &(1) \to b_i=0,\ \ h\,(0)=h\,(2)\approx 2/5 \\ &(2) \to b_i=1,\ \ h\,(1)\approx 1/5 \\ &(3),\ (4) \to b_i=2 \\ &\{b_i\} = \ldots,\ 222201220202001002112100,\ \ldots \end{aligned}$$

ACFs of sequences $\{b_i\}$ with symbol repetition rates defined by Eq. (4.72) also show a notable similarity to those of independent random signals [see Eq. (4.64)]. For the majority of shifts $s\Delta t$, the relationships

$$R_{ACF}(\tau)=R_\alpha=(\alpha_{PR}q^n-1)/(q^n-1) \quad \text{at} \quad 2k<n \tag{4.74}$$

$$\alpha_{PR}=\Big(\sum_j l_jB_j\Big)^2/q^k\sum_j l_jB_j^2 \tag{4.75}$$

hold if probabilities of symbols in Eq. (4.64) are equal to repetition rates of symbols in Eq. (4.72).

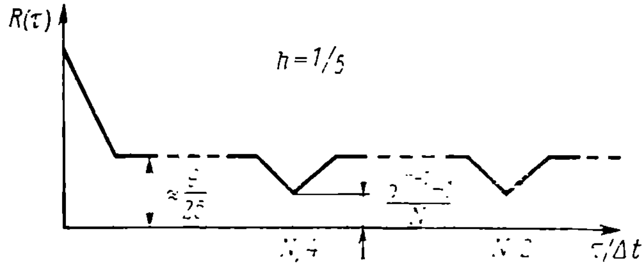

Fig. 4.8. ACF for a binary PNS

Generation of binary PNSs with specified repetition rates. This method for transforming the multisymbol M-sequences lends itself particularly for the production of binary PNSs with symbol repetition rates not equal to 1/2.

Binary PNSs with specified repetition rates of symbols can be had using generation and transformation of 3-, 5-, and 7-symbol M-sequences by binary logic circuits. For $k = 1$, maximally independent binary PNSs are obtained with symbol repetition rates of $h(\ldots) = 1/3, 2/3, 1/5, 2/5, 3/5, 1/7, 2/7 \ldots$

Example

Generating of a binary PNS with $h(+1) = 1/5$, $h(-1) = 4/5$. The 5-symbol M-sequence is transformed as follows:

$$\left.\begin{array}{l} q=5,\ q_T=2,\ k=1 \\ (0),\ (1),\ (2),\ (3) \rightarrow b_i = -1,\ (4) \rightarrow b_i = +1 \end{array}\right\} \quad (4.76)$$

$$R(st) = (9 \cdot 5^{n-2} - 1)/(5^n - 1) \approx 9/25 \quad \text{at} \quad s \neq j(5^n - 1)/4$$

Figure 4.8 depicts an ACF plotted according to Eq. (4.76). Safe for three points within the period, the ACF in Fig. 4.8 is essentially similar to ACF for a binary random signal with independent values at probabilities $p(+1) = 1/5$, $p(-1) = 4/5$.

Transformation with a digital filter. To seek special multisymbol PNSs, the transformation with a non-recursive digital filter may be used (Fig. 4.9). We suppose that $m < n$ and

$$b_i = \sum_{\nu=1}^{m} w_\nu a_{i-\nu} \quad (4.77)$$

If the sequence $\{a_i\}$ is obtained in a binary-coded form, $a_i = (x_{1,i}, \ldots, x_{e,i})$, $x_{i,j} \in \{0, 1\}$, then to split the coefficients w_ν additional degrees of freedom are available

$$b_i = \sum_{\nu=1}^{m} \sum_{u=1}^{e} w_{u,\nu} x_{u,i-\nu} \quad (4.78)$$

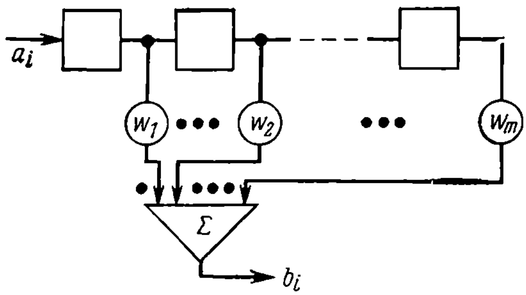

Fig. 4.9. Realization of weighted summation principle

Using the convolution of individual functions (ladder function with step height A_j and repetition rate h_j)

$$F_\nu = \sum_{j=0}^{q-1} h_j \delta(A - w_\nu A_j) \tag{4.79}$$

an equation is derived for the distribution of the sum F_b in the form of a polynomial in variable z

$$\mathscr{L}\{F_b\} = \prod_{\nu=1}^{m} \left(\sum_{j=0}^{q-1} h_j z^{w_\nu A_j}\right) \tag{4.80}$$

From Eq. (4.80) for the case of binary coding we have

$$\mathscr{L}\{F_b\} = \prod_{\nu=1}^{m} \left(\sum_{j=0}^{q-1} h_j z^{\sum\limits_{\mu} x_{\mu,\nu}(j) w_{\mu,\nu}}\right) \tag{4.81}$$

Example

$$q = 3, \quad a_i = (x_{1,i}, x_{2,i}), \quad x \in \{0, 1\}$$

is the ternary M-sequence in binary coding

$$w_{1,1} = w_{1,2} = w_{1,3} = w; \quad w_{2,\nu} = 0$$

$$\mathscr{L}\{F_b(z)\} = \prod_{\nu=1}^{3} \frac{1}{3}(z^0 + z^w + 2^0)$$

$$= \frac{1}{27}(2 + z^w)^3 = \frac{1}{27}(8 + 12z^w + 6z^{2w} + z^{3w})$$

The autocorrelation function of the sequence of output signal $\{b_i\}$ of the nonrecursive digital filter (see Fig. 4.9) may be sought as ACF of M-sequence. Simple expressions are derived for binary

and ternary M-sequences

$a_i \in \{+1, -1\}$:

$$R_b(s\Delta t) = \frac{1}{N} \begin{cases} 2^n \sum_{q=1}^{m} w_\nu^2 - \beta^2 & \text{at } s = 0 \\ 2^n \sum_{\nu=1}^{m-s} w_\nu w_{\nu+s} - \beta^2 & \text{at } s = 0, \ldots, \\ & m-1 \mod N \\ -\beta^2 \text{ otherwise} & \end{cases} \tag{4.82}$$

at $\beta = \sum_{\nu=1}^{m} w_\nu$

$a_i \in \{+1, 0, -1\}$:

$$R_b(s\Delta t) = \frac{1}{N} \begin{cases} \pm 2 \cdot 3^{n-1} \sum_{\nu=1}^{m} w_\nu^2 \\ \text{at } s = \begin{cases} 0 \\ N/2 \end{cases} \\ \pm 2 \cdot 3^{n-1} \sum_{\nu=1}^{m-s} w_\nu w_{\nu+s} \\ \text{at } s = \begin{cases} 1, \ldots, m-1 \\ N/2 - (1, \ldots, m-1) \mod N \end{cases} \\ 0 \quad \text{otherwise} \end{cases} \tag{4.83}$$

Concluding remarks. Two methods have been considered for the generation of PNSs with specified symbol repetition rates by using the generation of binary and multisymbol M-sequences: logical correlation of the output and weighted summation (nonrecursive digital filtration). These principles that are distinguished for simplicity of engineering realization might be extended to cover the combination of two or more PNSs. Such PNSs may be used as input signals for other methods of generating PNSs with a given distribution. In reference [434], signals with various distributions have been obtained by transformation of PNS time intervals into amplitudes, evaluation of these amplitudes, and reverse transformation into time intervals. In work [485], as a universal principle, the derivation is suggested of conditional probabilities through "conditional bit reading". Using the counting of a sufficiently large number

of binary PNSs elements and the digit-to-analog converting, PNSs can also be produced with a normal amplitude distribution [446].

Owing to rapid development of microprocessor technology, new prospects open up to obtain PNSs with a given statistical behavior.

4.5. Realization of Large-Base PNSs

General. Requirements placed upon large-base PNSs in communication systems are as follows:

relative simplicity of formation and treatment of PNS with 10^4-10^6 bases;

speedy re-adjustment of transmitting or receiving equipment for any signal of the used signal system with volume 10^4-10^8;

minor losses in shaping and processing of PNSs, as due to losses it is necessary to increase further the PNS base and broaden the bandwidth.

It follows from the above that in modern wideband communications it is advisable to adopt PSK, DF, and DCF signals. In principle, of course, PSK and DF signals may be obtained with bases 10^4-10^6, but to deal with them extremely elaborate apparatus will be required. Therefore, with bases $B = 10^4$-10^6 it is advantageous to utilize a DCF-PSK signal that shares the properties of PSK and DF signals. DCF-PSK signals enable these bases to be obtained with a practically acceptable number of frequency channels M and base N_0 of PSK signals. The base of DCF-PSK signals is

$$B = N_0 M^2 \tag{4.84}$$

The DCF-PSK signals show the following advantages:

they permit of large bases with acceptable number of frequence channels;

the discrete structure of DCF-PSK signal allows digital methods and equipment to be used widely in the formation and treatment of both DF structure and PSK signal:

the discrete structure of DCF-PSK signal and nonuniform distribution of its energy in the frequency-time plane reduces the probability of DCF-PSK signal corruption by the structural interference;

the availability of two qualitatively different signals (DF and PSK) in the DCF-PSK signal allows such a difference in rapid detection of the DCF-PSK signal and achievement of synchronization to be used;

the availability of M parallel channels improves the reliability of apparatus for the formation and treatment of large-base DCF-PSK signals, which is one of substantial requirements with such bases.

For these reasons we limit oneself with the issues of shaping and processing of DCF-PSK signals.

Shaping and processing of DCF-PSK signals. Algorithms of construction of large DCF-PSK systems are based on the algorithms of construction of PSK and DF signals and principally may be worked out through the determination of these latter. There is no asserting, however, that algorithms of construction of large systems of DCF-PSK signals are only controlled by those for PSK and DF signals, for in selecting the construction algorithms of DCF-PSK signals with bases 10^4-10^6 the methods of formation and treatment of these signals are to be taken into account, and also the methods of pulling into synchronism. Undoubtedly, the methods of technical realization are to exert the main influence on the choosing of algorithms of construction of DCF-PSK signals. We consider the peculiarities of shaping and processing such signals.

The coherent DCF-PSK signal is a signal for which initial phases of all frequency elements are related to each other by a determinated dependence, the phase commutation moments of PSK signals occur at zeros of the carrier oscillations, and the number of periods of the carrier oscillation in each frequency element is a multiple of the period of reference oscillation. Meeting the requirements of large-base PNS coherence is a complicated engineering problem.

The treatment of DCF-PSK signals is possible, using either a correlator (correlation treatment) or a matched filter (filter treatment) or both a correlator and a matched filter (correlation-filter treatment).

With the use of a correlator problems arise of forming the coherent DCF-PSK signal, providing linearity, large dynamic range of the integrator and rapid pulling into synchronism. The latter problem is the most difficult. The correlator necessitates a large time of pulling into synchronism. Based on the exhaustion method and prior uncertainty of the beginning of a PCF-PSK signal, it is necessary to pass on the average $B/2M$ information units, which might exceed the volume of the information transmitted. It is required therefore, first, to reduce the prior uncertainty as to the beginning of the DCF-PSK signal, and second, to use a multichannel (in time) correlator. If then in designing a communication system the prior uncertainty of the start of DCF-PSK signal cannot be reduced, then the correlation filter treatment methods should be adopted.

The matched filter for the DCF-PSK signal with bases 10^4-10^6 and large durations is hardly to be realized even in advanced electronic circuits and LSCs. At the same time, it will be remembered that ACF at the output of a matched filter enables the signal start to be segregated much easier, whereby providing a faster pulling into synchronism. It would seem therefore advisable to utilize a known correlation-filter treatment in two stages: (1) faster pulling into synchronism through the accumulation of ACF obtained from outputs of matched filters for parts of a DCF-PSK signal; (2) coherent treatment of the DCF-PSK signal a whole, using the correlator.

Faster pulling into synchronism. Synchrosequence (SS) will refer to the part of PSK signal which is handled by the matched filter. Depending on the number of SSs and their locations, various variants of circuit arrangement for faster pulling into synchronism are possible. A unit SS is practical (located in one arbitrary frequency element), two SSs (located in two frequency elements at intervals $T/2$), M SSs (located in each frequency element). In each of these cases the number of SSs can be increased n times. And ACFs will follow each other by batches with n in each.

According to the arrangement of SS, the following configurations are possible: at the end. beginning or at any interval of the frequency element, but identically for each and every frequency element. As far as noise immunity is concerned, all the configurations are in similar circumstances, but as regards the matched filters and correlator the first configuration may appear more advantageous.

For pulling into synchronism, matched filters alone are not enough. An accumulator is needed to sum up the ACFs. The coherent accumulator should have equivalent base

$$B_{ca} = BKM_{ss}^{-1} \tag{4.85}$$

where B is the base of DCF-PSK signal, K is the number of SSs summed up, M_{ss} is the number of SSs in one DCF-PSK signal. The complexity of the coherent accumulator is similar to the matched filter for the whole of the DCF-PSK signal. Therefore the incoherent accumulator alone may be utilized, i.e. after the matched filter an envelope detector should follow, then the ACF envelopes are summed up in the incoherent accumulator. With incoherent accumulation, the losses are the higher, the lower the signal-to-noise ratio at the detector input.

Handling of large-base PNSs. The pseudo-noise signals with bases 10^4-10^6 can be realized in communication systems with bandwidth from one to 10^3 MHz and signal duration from 10 μs to 1 s. The fabrication of passive matched filters with such ratings appears not feasible, which is testified by the data of advanced passive matched filters based on devices with surface acoustic waves (SAW) and on charge-coupled devices (CCD). At present, with relatively large durations of signals for passive matched filters, a base of 10^3 is attainable. It might be supposed that in the ensuing years passive matched filters with bases 10^4 will be developed, but hardly ever filters with bases 10^3-10^6.

Shown in Fig. 4.10 is a diagram [279] illustrating extreme characteristics of SAW, both being produced and still developed. Fig. 4.11 gives a diagram of extreme characteristics of CCD [6, 200]; peristaltic (PCCD), with a deep channel (DC), and with a surface channel (SC).

The principle of active matched filters—correlators with bases 10^4-10^6—is simpler than that of passive ones. To realize them, the

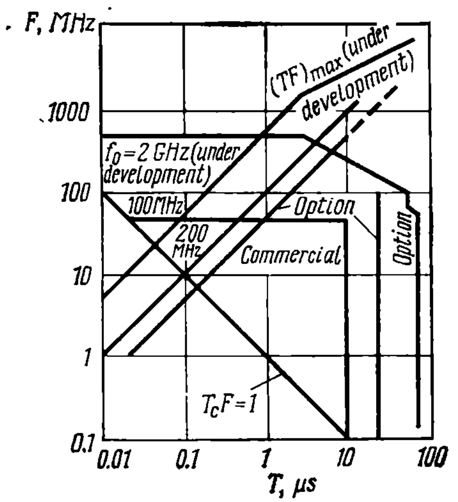

Fig. 4.10. Limiting characteristics of multitap delay lines

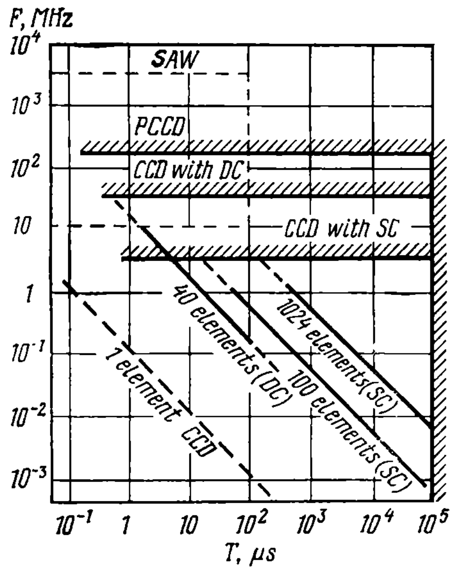

Fig. 4.11. Extreme characteristics of CCD and SAW-based devices

generator of reference signals (GRS), multiplier, and integrator are needed. The most sophisticated one with bases 10^4-10^6 is the generator of reference signals that forms a coherent PNS with such a base.

Figure 4.12 presents a popular block diagram of a receiver with a coincidence-type PNS treatment. Behaviour and the potentialities of realization of its components depend on the base, kind, duration, and width of PNS spectrum. Central to the diagram are the matched filter, synchronizer, and correlator. With 10^4-10^6 bases the realization of each of these devices is not trivial.

The matched filter should treat a part of PNSs in a coherent manner, and the synchronizer should effect incoherent accumulation. The creation of the coherent accumulator with so large a base and duration does not appear feasible at the present time. The principal problems here are: determination of the part of PNSs treated in a coherent manner, and selection of the matched filter; determination of the structure of the incoherent accumulator and of losses in such a method of pulling into synchronizm; determination of the methods of shaping of coherent correlators.

Discrete frequency signals. In this case the signal base is $B = M_1^2$, where M_1 is the number of unmodulated pulses in a signal (number of frequency elements or frequency channels). With the base $B = 10^6$ the number of channels (elements) is equal to $M_1 = 1\,000$. Accordingly, the relative time of search is given by

$$L_{sDF} = M_1 N_{syn} = \sqrt{B}\, N_{syn} \qquad (4.86)$$

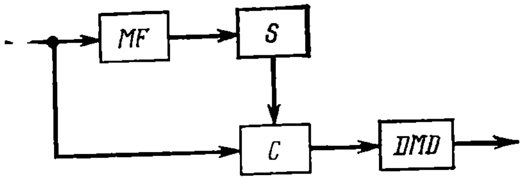

Fig. 4.12. Block diagram of receiver with integrated PNS processing: MF—matched filter; S—synchronizer; C—correlator; DMD—decision making device

where N_{syn} is the number of pulses accumulated in synchronizer. At $B = 10^6$, $M_1 = 10^3$, $N_{syn} = 5$, we have $L_{sDF} = 5\,000$, and at $N_{syn} = 10$ $L_{sDF} = 10\,000$. The comparison with PSK signals for a number of correlators of $K_{cor} = 400$ shows that the correlation search of DF signals using one correlator would result in a shorter search time than the search of PSK signals with a large number of correlators. If we assume that such a generator of DF signals might be developed, the creation of the K-channel correlator for the DF signal will be feasible. The relative search time will then be

$$L_{sDFk} = M_1 N_{syn}/K_{cor} = \sqrt{B}\, N_{syn}/K_{cor} \tag{4.87}$$

Composite signals. Consider first discrete composite frequency signals with phase-shift keying (DCF-PSK). Their base $B = B_0 M^2$, where B_0 is the base of PSK signal in one frequency element, M is the number of frequency elements. The total number of pulses of PSK signals in a DCF-PSK signal, $N_{PSK} = B_0 M$, is equal to the number of uncertainties as to the delay. The relative time of search

$$L_{sDCF\text{-}PSK} = B_0 M N_{syn} = B N_{syn}/M \tag{4.88}$$

i.e. M times shorter than for PSK signals. If the multichannel correlation treatment is used and the number of channels K_{cor} is set equal to M, then

$$L_{sDCF\text{-}PSKc} = B N_{syn}/M^2 \tag{4.89}$$

At $B = 10^6$, $M = 20$, $N_{syn} = 5$ the relative search time is $L_{sDCF\text{-}PSKc} = 12\,500$ and with $N_{syn} = 10$, $L_{sDCF\text{-}PSKc} = 25\,000$. The same search time is required for the PSK signal, but for the DCF-PSK signal 20 correlators are required, and for the PSK signal, 400 correlators.

We now consider discrete composite frequency signals with frequency-shift keying (DCF-FSK). Their base $B = M_1^2 M_2^2$, where M_1 is the number of external frequency elements, M_2 is the number of internal frequency elements in each external one. The total number of elements equals $M_1 M_2$. If we set $M_1 = M_2$, the base $B = M_2^4$, and the number of elements $N_{DF} = M_2^2$. Then the relative search

time will be given by

$$L_{sDCF\text{-}FSK} = M_2^2 N_{syn} = \sqrt{B}\, N_{syn} \tag{4.90}$$

which coincides with Eq. (4.86). With multichannel correlation treatment, assuming $K_{cor} = M_2$, we have

$$L_{sDCF\text{-}FSKc} = M_2 N_{syn} = \sqrt{B}\, N_{syn}/M_2 \tag{4.91}$$

which structurally corresponds to Eq. (4.87). At $B = 10^6$, $M_2 \approx 32$. If we set $N_{syn} = 5$, then by Eq. (4.91), $L_{sDCF-FSKc} = 158$; and at $N_{syn} = 10$, $L_{sDCF-FSKc} = 316$. It should be emphasized that for DCF-FSK signals the shortest relative search time has been obtained. If in Eq. (4.87) we put $K_{cor} = 32$, then the same results will hold for DF signals. But the structure of DCF-FSK signals is more suitable for multichannel correlation treatment.

Note that the shortest search time with correlation reception is associated with DCF-FSK and DF signals. The next are DCF-PSK signals, and the longest search time have the systems with PSK signals. At the same time, at $B = 10^5$-10^6 correlators are easier realized for DCF-PSK signals.

Synchronization. If a multichannel filter treatment is utilized, every frequency element, partly or wholly, is treated in a coherent manner using a matched filter. Suppose that each frequency element contains Q synchrosignals with duration T_{syn} each. Two arrangements of synchrosignals are possible: uniform and sequential. With the uniform arrangement, synchrosignals are spaced T_0/Q apart. In that case the repetition rate, or synchronization frequency, will be

$$F_{syn} = Q/T_0 = MQ/T$$

It should be noted that synchrosignals may be similar in form, but subject to additional phase modulation corresponding to the multiplication by $+1$ or -1 in the selected code sequence.

The maximum number of synchrosignals $Q_{max} = T_0/T_{syn}$ and the maximum repetition rate is $F_{r,\,max} = MQ_{max}/T = 1/T_{syn}$.

With the sequential arrangement, synchrosignals come in succession. They form a group (or packet) of signals. The repetition rate F_g of the group is always equal to $1/T_0 = M/T_0$.

The uniform arrangement singles out the synchronization frequency, but does not permit of simultaneous determination of the beginning of a frequency element. For this purpose the "review" of all the frequency elements is required. On the surface of it, the sequential arrangement of synchrosignals enables the frequency element start to be found. But the uncertainty of measurement of start will be compatible with the packet duration, as ACF of a packet of synchrosignals exhibits many peaks. Therefore the accuracy of measurement will be dictated by the "envelope" of ACF for the packet of synchropulses rather than by ACF of an individual synchrosignal.

We will concern ourselves now with the uniform arrangement of synchrosignals. By way of example, let us look at DF signals, $Q = 1$. The signal contains M_1 elements, where $M_1 = \sqrt{B}$. Denote by L_s the number of PNSs to be accumulated to have at the output of the incoherent accumulator the required S/N ratio h_s^2. The relationship has the form

$$h_s^2 = h_0^2 \gamma (h_0^2) M_1 L_s \tag{4.92}$$

where the losses are

$$\gamma (h_0^2) \approx h_0^2/2 \tag{4.93}$$

The S/N ratio at the output of an elementary (channel) matched filter for the DF signal is

$$h_0^2 = (P_s/P_{int})_{in} M_1 \tag{4.94}$$

Substituting Eq. (4.93) into Eq. (4.92) gives

$$h_s^2 = 0.5 (P_s/P_{int})_{in}^2 M_1^3 L_s \tag{4.95}$$

Denoting

$$A = (2h_s^2/B^2)(P_{int}/P_s)_{in}^2 \tag{4.96}$$

we arrive at

$$L_s = AM_1 = A\sqrt{B} \tag{4.97}$$

It follows from Eq. (4.97) that the number of accumulated NLSs is the smaller, the smaller the quantity A of Eq. (4.96), i.e. the larger the PNS base. The quantity A can be expressed otherwise considering that in the coherent treatment of the PNS as a whole the signal-to-noise ratio at the output is given by

$$h_c^2 = (P_s/P_{int})_{in} B \tag{4.98}$$

Substituting Eq. (4.98) into Eq. (4.96) gives

$$A = 2h_s^2/h_k^4 \tag{4.99}$$

The total relative search time is

$$L_{sDF} = L_s + M_1 N_{syn} \tag{4.100}$$

as it is necessary to "review" $M_1 = \sqrt{B}$ elements of uncertainty equal to the number of frequency elements. By "review" we mean the finding of an PNS in each and every uncertainty element, using a coherent correlator tuned to PNS in general, as it has been noted earlier. Substituting Eq. (4.97) into Eq. (4.100) and considering that $M_1 = \sqrt{B}$ gives the relative time of PNS search

$$L_{sDF} = (A + N_{syn})\sqrt{B} \tag{4.101}$$

At $B = 10^6$, $A = 10^{-2}$, $N_{syn} = 5$ we have $L_{sDF} = 5\,020$, and at $N_{syn} = 10$ $L_{sDF} = 10\,200$, which is rather significant. If a multichannel correlator is used with the number of channels $K_{cor} = \sqrt{M_1} = \sqrt[4]{B}$, then

$$L_{sDFc} = (A + N_{syn}/B^{1/4})\sqrt{B} \tag{4.102}$$

Using this technique, the characteristics of the pulling process into synchronism can be found for DCF-PSK, DCF-FSK, and PSK signals.

In case of PSK signals, the time of pulling into synchronism is the longest. In the correlation treatment, to reduce the characteristic time of search, it is more appropriate to make use of DF or DCF-FSK signals. The multichannel matched filter is inferior to the correlator in treatment of DF signal, but provides remarkable gain in treatment of DCF-PSK and DCF-FSK signals. The single-channel matched filter gives good results with DCF-PSK and DCF-FSK signals only. The best results are obtained when coherent treatment of the whole frequency element is used.

Chapter

5

Synthesis of Receivers under Prior Uncertainty

5.1. Problem Statement

Introductory remarks. In formulating the problems of synthesis it is assumed that some desired properties of the system are given, and it is required to find the structure of the system. Not infrequently the desired properties are given in the guige of a performance quality criterion. In that case the system synthesis is said to be optimal (in the sense of the specified quality criterion).The statistic synthesis is associated with the use of stochastic models of the input (signals, interference) and some parameters of the system. At all times uncertain situations arise which emphasise the information significance of stochastic models.

In the statistic synthesis of a system two kinds of uncertainty are distinguished. The synthesis is referred to as statistic synthesis with complete prior information if the distribution functions for input variables and additional constraints are known. Should some prior data be unknown, the synthesis is then said to be the statistic synthesis under prior uncertainty.

In optimal synthesis theory for systems with complete prior information, the Bayes approach is widely recognized, that is based on the minimum average risk criterion. But normally not all the prior data needed to employ the results of the Bayes theory are available to the designer of the system, therefore at present the theory of statistic synthesis is developed intensively with the view of devising methods of solving the problem of prior uncertainty to obtain structures immune to changes in the probability distribution function of input effects.

Two varieties of statistic synthesis under prior uncertainty. In the optimal algorithm of the system performance not all the data observed at the input are used, but only a minimum sufficient statistic representing one or more functions (or functionals) of the input data. As will be recalled, the minimal sufficient statistic for solving many

problems of system synthesis (detection, discrimination, classification and identification) is the likelihood ratio (scalar or vecto1) if observations are represented as a discrete sample or is a functional of likelihood ratio if observations are represented in an analog form without any sampling in time.

Under conditions of prior uncertainty, the computation of the likelihood ratio may appear impossible. In this connection two kinds of problems in statistical synthesis are distinguished: those with parametrical and non-parametrical prior uncertainty. In problems with parametrical prior uncertainty, the family of likelihood ratios (likelihood functionals) is parametric, i.e. has the form

$$l_n(\mathbf{x}|\boldsymbol{\vartheta}, \theta) = w_n(\mathbf{x}|\boldsymbol{\vartheta}, \theta)/w_n(\mathbf{x}|0, \theta); \tag{5.1}$$

$$l[x(t)|\boldsymbol{\vartheta}, \theta] = \lim_{n\to\infty} l_n(\mathbf{x}|\boldsymbol{\vartheta}, \theta) \text{ (in a regular case)} \tag{5.1a}$$

where $\mathbf{x}$ is the vector of sample values (observed sample), $\boldsymbol{\vartheta}$ is the vector parameter of the signal, θ is the vector of disturbing parameters (interference, unknown surroundings), n is the sample size, $x(t)$ is the continuous realization of the input process over the observation interval.

The form of the sample distribution function (likelihood function) in that case is assumed to be known.

In problems with non-parametric uncertainty, the family of likelihood ratios (likelihood functionals) is non-parametric, i.e. here the form of distribution functions of an interference (noise) and the mixture of signal and noise is unknown, and hence the class of likelihood ratios may be very wide (e.g. consist of continuous non-negative functions).

For estimation and filtering problems, in most cases the minimum sufficient statistic is the maximum posterior uncertainty density of the quantity at hand (generally, a vector one). And for this class of problems also under prior uncertainty conditions the above classification can be utilized.

Statistical synthesis requirements. Before we go on to consider the methods of dealing with prior uncertainty, we will indicate some of the basic requirements that may be placed upon the statistical synthesis of information systems, at least in the stage of preliminary design of these systems. This will permit us to conclude with a comparison of various methods of synthesis.

From rather general considerations it follows that synthesized algorithms of system operation should meet the following requirements (all or at least a part of them). Thus, the system should provide:

optimality as to a specific quality criterion for any finite size of sample (any finite observation time);

asymptotical optimality;

optimality with additional limitations on the kind of algorithm (e.g. after demodulation, digital, etc.);

stability for a wide range of distribution of signals and noise;

allowing for correlations and nonstationarity of processes;

analysis of performance characteristics of the algorithm at finite observation time.

Basic methods to overcome the prior uncertainty. If we embark on a major classification of works devoted to the statistical synthesis of systems under prior uncertainty, the following four principal methods involved may be distinguished: classical theory of statistical decision making, non-parametrical statistics technique, adaptive approach, and use of special principles of asymptotic optimality in non-adaptive and adaptive versions.

In what follows, these methods will be applied to the problems of signal detection in a noise background, which are formulated in terms of alternative statistical hypotheses testing. Not infrequently the above methods can be extended to include multialternative identification and classification of signals, and parameter estimation.

5.2. Methods of Classical Theory of Statistical Decisions

By way of example of a problem with parametrical prior uncertainty we will consider a traditional problem of composite hypotheses testing:

hypothesis H_0 (no signal): the likelihood function of $\mathbf{x}$ being observed is $w_n(\mathbf{x} \mid 0, \theta)$, $\theta \in \Theta_{\text{int}}$;

hypothesis H_1 (signal is present): the likelihood function of $\mathbf{x}$ being observed is $w_n(\mathbf{x} \mid \vartheta, \theta)$, $\vartheta \in \vartheta_s$, $\theta \in \Theta_{\text{int}}$.

Let $u(\mathbf{x})$ be sufficient statistic for the signal parameter ϑ, and $\mathbf{t}(\mathbf{x})$ is sufficient statistic for the disturbing parameter θ, and let the likelihood function in the formulated problem belong to an exponential family, so that it may be represented as

$$w_n(\mathbf{x} \mid \vartheta, \theta) = C(\vartheta, \theta) \exp(\vartheta u + \theta' \mathbf{t}) \tag{5.2}$$

Then the decision algorithm can be written as follows:

$$\Phi(u, \mathbf{t}) = \begin{cases} 1, \text{ signal is present if } u(\mathbf{x}) \geqslant c(\mathbf{t}) \\ 0, \text{ no signal if } u(\mathbf{x}) < c(\mathbf{t}) \end{cases} \tag{5.3}$$

It is an unbiased, uniformly most powerful (UMP) test of the null hypothesis H_0 versus the alternative hypothesis H_1 with constant value α of false alarm probability for all distributions belonging to the parametrical family $w_n(\mathbf{x} \mid 0, \theta)$, $\theta \in \Theta_{\text{int}}$.

$$\mathrm{E}\,\{\Phi(u, \mathbf{t}) \mid \mathbf{t}, H_0\} = \alpha \tag{5.4}$$

the threshold $c(\mathbf{t})$ in the algorithm of Eq. (5.3) should be defined by Eq. (5.4). But the use of Eq. (5.4) for these purposes gives rise to

many difficulties, and to overcome these the property of invariance of the rule of Eq. (5.3), as related to a group of observation, might be called into play.

To illustrate these points, we now consider the problem of detection of the signal $\mu s(t)$, $\mu > 0$, ($s(t)$ is known) in the presence of a nonstationary correlated additive normal noise with zero mean and correlation function $\sigma^2 R(t, y)$, the variance σ^2 characterizing the average power of the noise being unknown.

Using the orthogonal expansion of a normal random process, we obtain from observation in the interval $(0, T)$ of realization of $x(t)$ the independent samples

$$x_k = \sqrt{\lambda_k} \int_0^T x(t)\, \varphi_k(t)\, dt, \quad \text{for} \quad k = 1, \ldots, N \tag{5.5}$$

where λ_k and $\varphi_k(t)$ are eigenvalues and eigenfunctions of the integral equation

$$\varphi(t) = \lambda \int_0^T R(t, y)\, \varphi(y)\, dy \quad 0 \leqslant t \leqslant T \tag{5.5a}$$

Under hypothesis H_0 mean values of x_k are equal to zero, and for the alternative hypothesis H_1

$$\mathrm{E}\{x_k | H_1\} = \mu s_k = \mu \sqrt{\lambda_k} \int_0^T s(t)\, \varphi_k(t)\, dt \tag{5.6}$$

Variance x_k equals σ^2 under the both null-hypothesis and alternative. For the substitution of the finite sample $\mathbf{x} = (x_1, \ldots, x_N)$ for $x(t)$ not to be associated with a material loss of useful information, this requires that $N \gg 1$.

The sufficient statistic for the signal parameter will be

$$U(\mathbf{x}) = \mathbf{s}'\mathbf{x} \tag{5.7}$$

and for the disturbing σ^2

$$t(\mathbf{x}) = \sum_{k=1}^{N} x_k^2 \tag{5.8}$$

Rule (5.3) here is formulated as follows: the decision that the signal is present is made if

$$U(\mathbf{x}) \geqslant c\,[t(\mathbf{x})] \tag{5.9}$$

and that the signal is absent if the inequality inverse to Eq. (5.9) holds. To define the threshold function $c(t)$ pertinent to the given value α of the false alarm probability, use is made of the invariance of the rule of Eq. (5.9) with respect to a group of transformations that represents a multiplication of the realization observed by a positive constant. As a result, we arrive at a rule: signal is absent if

$$U(\mathbf{x}) \{[t(\mathbf{x}) - U^2(\mathbf{x})]/(N-1)\}^{-1/2} \geqslant t_\alpha \tag{5.10}$$

where t_α is the α-percentile of the Student distribution. Note that as $N \to \infty$, the quantity in braces converges in probability to σ^2. Proceeding from Eq. (5.10), we arrive, with $N \to \infty$, at a known optimal algorithm of detection in a problem with complete information on the interference parameter.

The advantage of the discussed principle in obviating the parametric prior uncertainty is the possibility of synthesis of UMP algorithms of detection with arbitrary finite sample sizes and also the possibility of analysis of performance characteristics of the detector. However, the scope of application of this principle is limited by the exponential family of distributions (additive normal noise) and by the invariance requirements for the decision algorithm with respect to a certain group of transformations.

A consistent treatment of similitude and invariance rules of decision making in testing composite hypotheses is given in the known book by Leman [426]. Minimax principles also employed to get over the prior uncertainty are discussed in a recent monograph by Repin and Tartakovsky [226]. Issues of surmounting the parametrical prior uncertainty by the methods of classical decision theory have been discussed in references [27, 28, 144, 156, 230].

5.3. Non-Parametrical Statistic Methods

Fundamentals of method. If an independent sample of observations is available, then non-parametric algorithms of signal detection can be synthesized, which are characterized by an invariable false alarm probability for any distributions of a stationary interference, this being valid with arbitrary finite sample sizes. Generally speaking, the probability of false dismissal will be higher than is theoretically possible and will be dependent on the interference distribution. The above non-parametric behavior is characteristic for sign, rank, and sign-rank algorithms of signal detection, in which the initial operation is the transformation of the observation vector $\mathbf{x} = (x_1, \ldots, x_n)$ accordingly into the vector of signs

$$\operatorname{sgn} \mathbf{x} = (\operatorname{sgn} x_1, \ldots, \operatorname{sgn} x_n),$$
$$\operatorname{sgn} x_k = x_k/|x_k| = 2u(x_k) - 1$$
$$u(x_k) = \begin{cases} 1, & x_k \geqslant 0 \\ 0, & x_k < 0 \end{cases} \tag{5.11}$$

the rank vector is

$$\mathbf{R}(\mathbf{x}) = (R_1, \ldots, R_n),$$
$$x_k = x^{(R_k)}, \quad x^{(i)} \leqslant x^{(j)}, \quad i < j \tag{5.12}$$
$$R = \sum_{k=1}^{n} u(x_i - x_k) = \frac{1}{2} \sum_{k=1}^{n} \operatorname{sgn}(x_i - x_k) + \frac{n}{2} \tag{5.12a}$$

and the vector of positive ranks

$$\mathbf{R}^+(\mathbf{x}) = (R_1^+, \ldots, R_n^+), \quad x_i = |x|^{R_i^+} \operatorname{sgn} x_i \tag{5.13}$$

If the distribution of stationary interference is symmetrical about zero, then the numbers with positive and negative signs in the independent sample of interference feature equal probability. As a positive value of signal appears, the probability of positive signs in the sample becomes larger than that for negative signs, which makes the detection of signal possible. Thus, the sign detection algorithms respond to a signal in the presence of an arbitrary interference distributed symmetrically about zero (zero mean).

For a homogeneous independent sample all the rank vectors show equal probabilities. As the total number of rank vectors for a sample with size n is equal to $n!$, then

$$\mathrm{P}\{R = r_i \mid H_0\} = 1/n! \tag{5.14}$$

Thus, the utilization of the rank algorithm of detection retains the invariability of probability of false alarms for a stationary interference with arbitrary distribution. Moreover, the rank vector of stationary interference is invariant to memoryless transformation of a sample. Therefore the rank algorithm of detection retains its non-parametric property after a non-linear memoryless transformation of observed data.

As a rule, the quality of non-parametric algorithms is characterizable asymptotically only, when the size of a sample (observation time) grows infinitely. A quality criterion is the asymptotic relative efficiency (ARE) to be defined as follows.

Let the values of probability of erroneous decisions be fixed, i.e. those of false alarm α and false dismissal β. Let δ_{n_k} and $\delta^*_{n^*_k}$ be two sequences of detection algorithms to be characterized by the said probabilities of erroneous decisions for sample sizes n_k and n_k^*, respectively, and for $n_k \to \infty$, $n_k^* \to \infty$, when $k \to \infty$. Then the ARE of sequence $\delta^*_{n^*_k}$ as related to the sequence δ_{n_k} is equal, by definition, to

$$\rho = \lim_{k \to \infty} (n_k / n_k^*) \tag{5.15}$$

In its limiting form, ARE indicates that of the two detection algorithms being compared the more efficient is one for which, the probability of false alarms being given, a specified probability of correct detection is reached with smaller samples.

We now consider several examples of non-parametric detection algorithms practically realizable without any significant difficulty.

Sign algorithm. The sign algorithm of detection of the determinate signal $\mu s(t)$, $\mu > 0$, in the presence of an additive stationary inde-

pendent interference with zero mean and symmetric distribution, claims that a signal is present if

$$\sum_{k=1}^{n} u(s_k x_k) \geqslant c, \; s_k = s(t_k), \;\; x_k = x(t_k) \tag{5.16}$$

The threshold c is governed by a predetermined false alarm probability. The block diagram of such a detector involves a multiplier, ideal delimiter, adder, and threshold-comparison device.

It can be shown that the ARE of algorithm of Eq. (5.16) with reference to the linear algorithm

$$\sum\nolimits_{k=1}^{n} s_k x_k \geqslant c \tag{5.17}$$

optimal at any n from the Neyman-Pearson criterion for a normal interference, is equal to

$$\rho = 4\sigma^2 w_1^2(0)\, V_s^2 \tag{5.18}$$

where $w_1(x)$ and σ^2 are the probability density and variance of interference, respectively:

$$V_s^2 = \lim_{n\to\infty} a_{|s|,\,n}^2 / W_{s,\,n} \tag{5.19}$$

$$a_{|s|,\,n} = \frac{1}{n}\sum_{k=1}^{n} |s_k|, \quad W_{s,\,n} = \frac{1}{n}\sum_{k=1}^{n} s_k^2 \tag{5.19a}$$

For a constant signal $V_s = 1$, and we arrive, from Eq. (5.18), at the known formula

$$\rho = 4\sigma^2 w_1^2(0) \tag{5.20}$$

Whence follows that for the normal distribution of interference $\rho = 2/\pi \approx 0.65$, and for the Laplace distribution $\rho = 2$, i.e. the use of sign algorithm with normal interference gives rise to a reduction of about 35% as compared with the optimal case, and with the Laplace interference the sign algorithm is twice as good as the linear one.

Rank algorithm. In general, rank non-parametric detectors are more efficient than sign ones. So, in the problem worked above, ARE of the sign-rank algorithm

$$\sum\nolimits_{k=1}^{n} s_k R_k^+ u(x_k) \geqslant c \tag{5.21}$$

with reference to the linear algorithm of Eq. (5.17) is

$$\rho = 12\sigma^2 \left(\int_{-\infty}^{\infty} w_1^2(y)\, dy \right)^2 a_s^4 / W_s \tag{5.22}$$

where

$$a_s = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} s_k, \quad W_s = \lim_{n\to\infty} W_{sn} \tag{5.22a}$$

For a constant signal $a^2/_sW_s = 1$, and Eq. (5.22) coincides with the formula known from the literature. In that case for a normal distribution of interference $\rho = 3/\pi \approx 0.955$, i.e. the algorithm of Eq. (5.21) is essentially as efficient as is the linear algorithm of Eq. (5.17).

If $a_S = 0$, i.e. the bias of the signal is zero (as is the case, for instance, with microwave signals), then $\rho = 0$. Here the rank algorithm is more appropriate. According to the simplest linear rank algorithm, the decision that the signal is present is made if

$$\sum_{k=1}^{n} s_k R_k \geqslant c \tag{5.23}$$

The ARE of this algorithm as related to the linear one of Eq. (5.17) will be

$$\rho = 12\sigma^2 \left(\int_{-\infty}^{\infty} w_1^2(y)\, dy \right)^2 (W_s - a_s^2) \tag{5.24}$$

The quantity ρ defined by Eq. (5.24) reaches its maximum value at zero bias of the signal.

Higher efficiency of rank algorithms as compared with sign algorithms is, of course, due to the more sophisticated detector, as the ranking of a sample calls for the number of operations equal to the square of the sample size.

Diversity reception. For the last example of utilization of non-parametric algorithms, we shall take the problem of detection of stochastic (fluctuating) signal. For this purpose at times a two-channel (diversity reception) system is adopted. When there is no signal, each channel has only an interference, and the processes occurring in the channels are independent. When a signal appears in both channels, a statistical coupling takes place. The optimal—by the Neyman-Pearson criterion—rule for the normal signal and normal interference is to be formulated as follows: the signal is present if

$$\sum_{i=1}^{n} (x_i + y_i)^2 \geqslant c \tag{5.25}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is an independent sample observed at the output of the first channel, and $\mathbf{y} = (y_1, \ldots, y_n)$ is an independent sample observed at the output of the second channel. It is supposed here that the signal and interference have zero mean and differ by their variances.

If probability densities for the signal and interference are arbitrary functions symmetrical about zero, the correlator of polarity coincidence might be used, which is essentially a non-parametric two-channel sign detector. A decision that a signal is present is made if

$$\sum_{k=1}^{n} \operatorname{sgn} x_k \operatorname{sgn} y_k \geqslant c \tag{5.26}$$

Relative to the optimum algorithm of Eq. (5.25), ARE of the algorithm of Eq. (5.26) for a normal signal and interference is given

by

$$\rho = 2w_1^4(0)(m_4 + \sigma^4) \tag{5.27}$$

where m_4, σ^2 and w_1 are the fourth moment, variance, and probability density of the interference, respectively.

Concluding remarks. It follows from the above that the major advantage of sign and rank algorithms of detection lies in the maintenance of a constant level of false alarms for a rather broad class of interference at any sample size. This non-parametric feature of algorithms, however, occurs at stringent requirements of independence of observations and stationary nature of the interference. In addition, the selection of one or another non-parametric algorithm is not based on any optimality criterion, but is done purely heuristically. The comparison of algorithms by using ARE is not always conducive to a motivated judgement as to the quality of algorithm at finite sample sizes. The smallness of ARE does not signify yet that a non-parametric algorithm must be inefficient with small samples.

The rank test theory with its applications is consistently discussed in monographs by Hajek and Sidak [380]. Tarasenko [257], in the book 'Communication Theory' [315], and in the third volume of Levin's monograph [164].

5.4. Adaptive Approach

To overcome prior uncertainty, use can be made of estimates, obtained by learning (with a supervisor or non-supervised), of distribution functions, their parameters or any other characteristics (e.g. posterior risk). Algorithms using supervising samples (classified when learnt with a supervisor or unclassified if non-supervised) are said to be adaptive.

The general quality criterion of an adaptive algorithm is its convergence in the unlimited growth of the size of a teaching sample to a suitable optimum algorithm obtained under conditions of complete prior information. Adaptive algorithms meeting this criterion are known as consistent. To be sure, not every adaptive algorithm is consistent, as any learning with samples introduces an additional uncertainty.

In problems with parametric prior uncertainty (e.g. in those with disturbing parameters), adaptive algorithms are produced from optimum ones by replacing unknown parameters with their estimates obtained by teaching samples. But it is to be noted that the consistency criterion of the adaptive algorithm does not yet uniquely determines the kind of estimates of unknown parameters. In practice such estimates are often those of maximal likelihood. In problems with non-parametric prior uncertainty, adaptive algorithms are

synthesized, using the estimates of unknown distribution functions (or functionals of these functions) constructed with the aid of teaching samples.

The choice of appropriate estimates of parameters or functions used in the construction of a consistent algorithm predetermines the rate of convergence of this adaptive algorithm. The acquisition of proper analytical estimates of the convergence rate constitutes a difficult and at times insoluble mathematical problem to be replaced by extensive computer-aided statistical experiments.

As a simplest illustration of the adaptive approach we take the problem of detection of a determinated signal $s(t)$ with a background of an additive uncorrelated normal interference with zero mean and unknown variance (average power). This is a simplified version of the problem considered above.

We first assume that as a result of learning with a supervisor a classified independent sample of the interference $\mathbf{y} = (y_1, \ldots, y_m)$ has been obtained. From this sample the estimate of maximum likelihood of the unknown variance of the interference is made

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{m} \sum_{i=1}^{m} y_i^2 \tag{5.28}$$

The sufficient statistic with known interference variance σ^2 in the problem in hand is $\mathbf{s}'\mathbf{x}/\sigma^2$, where $\mathbf{s} = (s_1, \ldots, s_n)$ is the vector of signal values; $\mathbf{x} = (x_1, \ldots, x_n)$ is the observed sample. Substituting for the unknown variance σ^2 its maximum likelihood estimate of Eq. (5.28), we arrive at an adaptive detection algorithm: the hypothesis that the signal is present is accepted if

$$\sum_{i=1}^{n} s_i x_i \Big/ \sum_{i=1}^{m} y_i^2 \geqslant c \tag{5.29}$$

In non-supervized learning, the maximum likelihood estimate of the unknown interference variance is derived from observed (information) sample under both null hypothesis H_0 (no signal), and alternative hypothesis H_1 (signal is present). In that case the estimate of likelihood ratio will be

$$\hat{l}(\mathbf{x}) = \max_{s,\,\sigma^2} W_n(\mathbf{x} \mid H_1) / \max_{\sigma^2} W_n(\mathbf{x} \mid H_0) = \hat{\sigma}^2_{H_0} / \hat{\sigma}^2_{H_1} \tag{5.30}$$

where

$$\hat{\sigma}^2_{H_0} = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \tag{5.31}$$

$$\hat{\sigma}^2_{H_1} = \frac{1}{n} \sum_{i=1}^{n} (x_i - s_i)^2 \tag{5.32}$$

The adaptive algorithm here dictates to adopt the hypothesis of the signal presence if

$$\sum_{i=1}^{n} x_i^2 / \sum_{i=1}^{n} (x_i - s_i)^2 \geqslant c \tag{5.33}$$

The adaptive approach is a universal means of circumventing prior uncertainty. With this approach, the above limitations inherent in the methods of classical decision theory and non-parametric statistic are absent. But the adaptive technique also is not without disadvantages. Extra time is required for learning, estimates of unknown characteristics are not to be derived uniquely with the result that algorithms proper are constructed on a heuristic basis, difficulties arise in estimating the convergence rate of the adaptive algorithm to the optimal one.

Extensive literature deals with the adaptive approach and synthesis of information systems under the conditions of prior uncertainty. Mention should be made of monographs by Tsypkin [283], Bellman [321], Petrov et al. [208], Repin and. Tartakovsky [226], and also articles [217, 228]. Fruitfulness of the adaptive approach to the solution of problems of synthesis is illustrated in Sec. 5.7 for the case of signal synchronization in communication systems.

5.5. Special Asymptotically Optimal Principle

Fundamentals of method. Both in the use of non-parametric statistics, and in the adaptation, the decision algorithm is constructed at finite sample sizes, and the quality criterion of the algorithm is asymptotic. In the first case this criterion is the ARE value; in the second, the consistency of the algorithm.

Unlike those approaches, let us concern ourselves directly with the asymptotic method for the construction of algorithm of noisy signal detection. The utilization of such asymptotically optimal (AO) algorithm in the limiting case appears of value with its fair convergence rate to the optimal one, and the resulting quality is not inferior, and possibly even superior, to algorithms derived proceeding from other principles.

Let $\lambda s(t)$, $\lambda > 0$ be a determinated signal and $\mathbf{x} = (x_1, \ldots, x_n)$ be an observed sample of size n. If we fix λ and increase n unlimitedly, then for the consistent detection algorithm with a fixed false-alarm probability the false dismissal probability tends to zero. On the other hand, if we fix n and make the signal amplitude, λ, approach zero, then for the same conditions the false dismissal probability will tend to unity. A case is of interest when simultaneously $\lambda \to 0$ and $n \to \infty$. It is found here that with a certain relation between λ and n there occurs a limiting value of the false dismissal

probability other than zero and unity. To have this requires that

$$\lambda \sqrt{n} = \gamma, \quad 0 < \gamma < \infty \tag{5.34}$$

The condition of Eq. (5.34) admits of a simple explanation. As the quantity $\lambda^2 n$ is proportional to the signal-to-noise ratio, this condition means a limitation to the signal-to-noise ratio in the limit.

Let $\lambda_1, \ldots, \lambda_n$ be a sequence of signal amplitude tending to zero at $n \to \infty$. Denote by $\beta_n(\delta_n, \lambda_n)$ the false dismissal probability with amplitude λ_n, when the algorithm of detection, δ_n, from a sample of size n is used. We will call the sequence δ_n^0 of detection algorithms asymptotically optimal, if for any other algorithm sequence δ_n at a fixed false-alarm probability

$$\lim_{n \to \infty} [\beta(\delta_n, \lambda_n) - \beta(\delta_n^0, \lambda_n)] \geqslant 0 \tag{5.35}$$

Here $\lambda_n = \gamma/\sqrt{n}$ and there exists a limit

$$\lim_{n \to \infty} \beta_n(\delta_n^0, \gamma/\sqrt{n}) = \beta(\delta^0, \gamma) \tag{5.36}$$

Interference with independent values. Suppose first that a determinated signal may appear in the presence of an arbitrary interference with independent values. In that case the sufficient statistic (likelihood ratio logarithm) is given by

$$\ln l(\mathbf{x}) = \sum_{i=1}^{n} \ln \frac{w(x_i \mid \lambda s_i)}{w(x_i \mid 0)}, \tag{5.37}$$

$$x_i = x(t_i), \quad s_i = s(t_i)$$

We will represent the probability density $w(x \mid \vartheta)$ in the form

$$w(x \mid \vartheta) = w(x \mid 0)\,[1 + \vartheta f(x) + \chi^2(x, \vartheta)] \tag{5.38}$$

where $\chi^2(x, \vartheta) \to 0$ in probability as $n \to \infty$ both under null hypothesis H_0, and under alternative hypothesis H_1. The function $f(x)$ will be

$$f(x) = \frac{\partial}{\partial \vartheta} \ln w(x \mid \vartheta)\,|_{\vartheta=0} \tag{5.39}$$

Using Eq. (5.38), it might be proved that both under H_0 and under H_1, an asymptotic equivalence of statistics is bound to take place on either side of the following expression

$$\ln l(\mathbf{x}) \sim \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i) + c, \quad \gamma = \lambda \sqrt{n} \tag{5.40}$$

The asymptotic equivalence signifies that for $n \to \infty$ the statistic on the right side of Eq. (5.40) converges in probability to the statistic on the left side.

The statistic

$$y_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f(x_i) \tag{5.41}$$

on the strength of the Central Limit Theorem under the condition that

$$\lim_{n\to\infty} \max_{0\leqslant i\leqslant n} s_i^2 \Big/ \sum_{i=1}^{n} s_i^2 = 0 \tag{5.42}$$

which holds practicaliy at all times, is asymptotically normal.

The asymptotically optimal algorithm of signal detection consists in comparing the statistic of Eq. (5.41) with a threshold dependent on a fixed false-alarm probability. To determine the threshold and probability of correct signal detection, it is sufficient to know the mean and the variance of the statistic of Eq. (5.41). These quantities respectively, are $(0, I_f W_s)$ under null hypothesis H_0, and $(\gamma I_f W_s, I_f W_s)$ under alternative hypothesis H_1

$$I_f = \mathrm{E}\{f^2(x) \mid H_0\} \tag{5.43}$$

$$W_s = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} s_i^2 \tag{5.43a}$$

In the case under discussion the detection device has a very simple structure. At the detector input there is a memoryless non-linear converter with a characteristic $f(x)$, followed by a correlometer. If the interference is a normal, additive one, then $f(x) = x/\sigma^2$, and the treatment algorithm will be linear. In the case of the Laplace additive interference

$$w(x \mid 0) = (a/2) \exp(-a \mid x \mid), \quad a > 0 \tag{5.44}$$

and the characteristic of the non-linear converter has the form

$$f(x) = a \operatorname{sgn} x \tag{5.45}$$

i.e. it represents the characteristic of the ideal delimiter. In such a case, the asymptotically optimal algorithm of detection coincides with the sign algorithm.

With independent samples, the theory of AO algorithms of detection permits of a dramatic extension of its scope of application.

Rank AO algorithm. The asymptotically optimal algorithm can be made non-parametric if the observed sample is subjected to a preliminary ranking. The non-parametric AO algorithm of detection is reduced to a comparison with the threshold of the statistic

$$z_n(R) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f\left[F^{-1}\left(\frac{R_i}{n+1}\right)\right] \tag{5.46}$$

where $F^{-1}(u)$ is the inverse of the interference distribution, and $f(x)$ is to be derived from Eq. (5.39).

Digital AO algorithm. Moreover, digital AO detection algorithms have been derived that use sampled data, quantized both in time and in amplitude. Here, compared with the threshold is the statistic

$$y_{nz}(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i f_z(x_i) \tag{5.47}$$

where

$$f_z(x) = \sum_{k=1}^{n} a_k \chi_k(x) \tag{5.48}$$

$$a_k = \frac{\partial}{\partial \vartheta} \ln p_k(\vartheta) \, |_{\vartheta=0} \tag{5.49}$$

$$p_k(\vartheta) = \int_{z_k}^{z_{k-1}} w(x \mid \vartheta)\, dx, \quad k = 1, \ldots, m \tag{5.50}$$

$$\chi_k(x) = \begin{cases} 1, & x \in E_k \\ 0, & x \bar{\in} E_k, \quad E_k = (z_k, z_{k+1}) \end{cases} \tag{5.51}$$

A number of AO algorithms solve the problem of detection of narrow-band signals with additional limitations peculiar to these signals; these include quadrature treatment, amplitude method, and phase method in its two versions (analog-discrete and digital).

Correlated interference. The theory of AO algorithms covers also the case of correlated interferences simulated by a sufficiently general model of the multi-connected Markov chain. The AO algorithm of detection is reduced here to comparing with the threshold of statistic (Fig. 5.1)

$$Y_n(\mathbf{x}_{-k}^{n}) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n} \mathbf{f}'(\mathbf{x}_{i-k}^{i})\, \mathbf{s}_{i-k}^{i} \tag{5.52}$$

where the components of the vector $\mathbf{f}$ are

$$f_j(\mathbf{x}_{i-k}^{i}) = \frac{\partial}{\partial s_{i-j}} \ln w\left(x_i \,\middle|\, \mathbf{x}_{i-k}^{i-1}, \ \mathbf{s}_{i-k}^{i}\gamma/\sqrt{n}\right)\Big|_{s=0} \tag{5.53}$$

$$\mathbf{x}_{i-k}^{i} = (x_{i-k}, \ldots, x_i), \quad \mathbf{s}_{i-k}^{i} = (s_{i-k}, \ldots, s_i) \tag{5.54}$$

In all the discussed above formulas, the statistics z_n, y_{nz}, Y_n are asymptotically normal, and the parameters of this limiting normal distribution are defined by sufficiently simple relationships.

General characteristic of AO algorithms. The AO algorithms considered above feature a certain structural stability, and their utilization requires a knowledge of the distribution of interference (arbi-

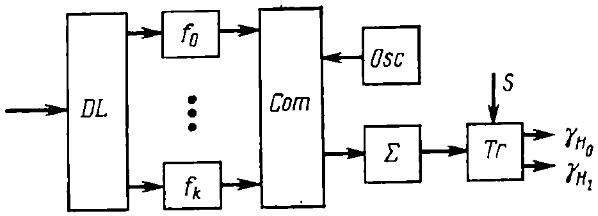

Fig. 5.1. Block diagram of AO detector

trary) and the way of its interaction with the signal. If a teaching sample of interference is available, then by formulating the estimate of characteristics of non-linearity and threshold, the AO adaptive detection algorithm can be derived to be employed under conditions of prior uncertainty of interference.

The theory of AO detection algorithm allows of taking a fresh look at the value of the Bayes theory for the model of a normal additive interference. Apart from their direct purpose, the results of this theory lend themselves to be used in the synthesis of AO detection algorithms whose realization involves non-linear transformation, with memory, of sufficient statistics exhibiting an asymptotically normal distribution. This enables the configuration of AO detector to be represented in the form of two devices. The first forms asymptotically sufficient statistics from prior data available, or using teaching samples. The second device is always a correlator.

The application of AO algorithms in a sublimiting case is effective at a sufficient convergence rate. It is important therefore to have adequate estimates of the number of samples (observation time) so that the use of AO algorithms does not lead to a relative loss of potential quality of the algorithm, this loss may be higher than a preset value. The analytical solution to this problem is very hard to derive. But extensive results have already been accumulated of computer-aided statistic simulation, which invite a measure of optimism. A relative loss of several percent is involved with sample sizes of the order of hundreds.

The derivation of AO algorithm theory is given in monograph [164] reviewing and systematizing the results of original works [158-160, 168-176, 427, 428].

5.6. Tradeoff of Considered Approaches to Synthesis under Prior Uncertainty

Table 5.1 characterizes the methods discussed above for overcoming prior uncertainty to meet the requirements of Sec. 5.1 for the statistical design of information systems. The "+" sign corresponds to

TABLE 5.1

Approach	Requirements					
	Optimality at any n	Asymptotical optimality	Optimality with additional limitations	Stability	Account of correlation and non-stationarity	Analysis of performance at any n
Classical decision theory	+	+	—	—	+	+
Non-parametric statistic	—	—	—	+	—	—
Adaptive approach	—	+	+	+	+	—
Special AO principle	—	+	+	+	+	—

satisfying the requirement, and the "—" sign signifies the failure to comply with the requirement.

The scrutiny of Table 5.1 enables the following conclusions to be drawn:

1. The most advantageous are the adaptive approach and special asymptotically optimal principle.

2. The non-parametric approach in its purest seems discouraging but when combined with other principles (e.g., AO) it shows promise.

3. In special cases (e.g., with a normal additive interference) the most advantageous is the approach associated with the classical solution theory.

5.7. Adaptive Synchronization in Communications

Introductory remarks. The process synchronization problem arises in various fields of technology. It is of paramount importance in television, radar, radio navigation, coherent communications, high-fidelity magnetic recording, and so forth. As applied to communication, synchronization consists in timing the periodic (quasiperiodic) processes at the send and receive ends of a communication system.

Statistical communication theory treats the synchronization problem on the basis of the theory of optimal reception. Besides the information parameters of signal vital for the transmission of basic information, non-information parameters are singled out which describe time relations in the signal and referred to in what follows as synchroparameters. Optimization problems are solved using information parameters. In the process, the problems of synchronization, i.e. estimation of synchroparameters, are dealt with either individually or in combination with estimation of information parameters [4, 239, 261, 496]. In the first case, the presence of an individual syn-

chrosignal is assumed, and then the synchronization problem is a usual problem of parameter estimation. The second case constitutes the problem of application of information signal in estimating the information- and synchroparameters, the latter problem being more important in present-day communication systems.

More often than not, the user is interested in reception quality as related to the information parameter alone. And of interest here are the following questions: how a synchronization device appears in the optimal (or close to optimal) receiver configuration, what estimates of synchroparameter make this device a reality, how the quality of algorithms is to be estimated. In this connection, here is considered the contribution of synchronization devices to the general reception algorithm, using various approaches to the reception problem. This difference is thought of as being due to the completeness of prior data available as related to the parameters of signal and loss function. The utilization is also discussed of the Bayes procedures and adaptive reception in synthesizing the near-optimal algorithms. It is shown that the synchronization system, as an independent structural element of the receiver, not always can be distinguished in a general reception problem, this being mostly conditioned by the problem statement, prior data employed for all the signal parameters, and solution technique.

Synchronization with Bayes reception procedure. Let the Markov process (θ, λ, y) be given. The component $y = y(t)$ is the result of observation in the presence of additive normal white noise with spectral density N of the signal $s(\theta, \lambda, t)$ depending on $\theta(t)$ of the information parameter and $\lambda(t)$-synchroparameters. We first consider a case where θ, λ are random processes with completely specified prior differential operators $\mathbf{L}_\theta$, $\mathbf{L}_\lambda$ and initial distribution densities $p_\theta(\theta)$, $p_\lambda(\lambda)$. For a given loss function $C(\theta, \hat{\theta})$ it is required to find the Bayes estimate $\hat{\theta} = \hat{\theta}(t)$ of the parameter θ at any instant of time t. The problem is tackled by using the equation of posterior probability density $w_t(\theta, \lambda)$ for a composite Markov process (θ, λ) expressed in the symmetrical form [254]

$$\frac{\partial}{\partial t} w_t(\theta, \lambda) = \mathbf{L} w_t(\theta, \lambda) + w_t(\theta, \lambda)[F(y, \theta, \lambda) - \overline{F}] \qquad (5.55)$$

where $\mathbf{L} = \mathbf{L}_\theta + \mathbf{L}_\lambda$ is the differential operator of the process (θ, λ);

$$F(y, \theta, \lambda) = (1/N)\, ys(\theta, \lambda, t) - (1/2N)\, s^2(\theta, \lambda, t)$$

$$\overline{F} = \int F(y, \theta, \lambda)\, w_t(\theta, \lambda)\, d\theta\, d\lambda$$

From Eq. (5.55) the equation of distribution density $w_t(\theta)$ can be derived for the information parameter

$$\frac{\partial}{\partial t} w_t(\theta) = \mathbf{L}_\theta w_t(\theta) + w_t(\theta)\,[\langle F\rangle(\theta) - \bar{F}] \tag{5.56}$$

where

$$\langle F\rangle(\theta) = \int_{\{\lambda\}} F(y,\ \theta,\ \lambda)\, w_t(\lambda \mid \theta)\, d\lambda$$

and $w_t(\lambda \mid \theta)$ satisfies the equation if θ belongs to a discrete finite set

$$\frac{\partial}{\partial t} w_t(\lambda \mid \theta) = \mathbf{L}_\lambda w_t(\lambda \mid \theta) + w_t(\lambda \mid \theta)\,[F(y,\ \theta,\ \lambda) - \langle F\rangle(\theta)]$$

with the initial conditions

$$w_t(\theta)|_{t=0} = p_\theta(\theta), \quad w_t(\lambda \mid \theta)_{t=0} = p_\lambda(\lambda).$$

The solution to Eq. (5.56) is used to calculate the conditional risk

$$R(\hat{\theta}) = \int_{\{\theta\}} C(\theta,\ \hat{\theta})\, w_t(\theta)\, d\theta \tag{5.57}$$

and the minimization of the latter defines the Bayesian procedure of estimation of the information parameter at complete prior specification of information parameter and synchroparameter.

In the transition to quasilinear algorithm of reception, the algorithms can be simplified through the transition to finite simultaneous differential equations for parameters of posterior distribution [254].

The incompleteness of prior data as to signal parameters, in particular about a synchroparameter, may be obviated with a parametric representation of the operator $\mathbf{L}_\lambda = \mathbf{L}_\lambda(\alpha)$ of the process $\lambda = \lambda(t)$ with a specified initial distribution of the parameter $p_\alpha(\alpha)$. In certain problems of signal reception such a description of a synchroparameter is typical. The statement of reception problem here consists in the following. A Markov process $(\theta,\ \lambda,\ y)$ is given, whose components θ, λ are defined by the operators $\mathbf{L}_\theta$, $\mathbf{L}_\lambda(\alpha)$. Then a loss function $C(\theta,\ \hat{\theta})$ is introduced. It is required to find the estimate $\hat{\theta}$ meeting the condition

$$\min_{\hat{\theta}} R(\hat{\theta} \mid y) = \min_{\hat{\theta}} \int C(\theta,\ \hat{\theta}(t))\, w_t(\theta)\, d\theta \tag{5.58}$$

Proceeding from the equation of posterior probability density for extended process $(\theta,\ \lambda,\ \alpha)$ we arrive at the equation

$$\frac{\partial}{\partial t} w_t(\theta) = \mathbf{L}_\theta w_t(\theta) + w_t(\theta)\,[\langle F\rangle(\theta) - \bar{F}] \tag{5.59}$$

where

$$\langle F\rangle(\theta)=\int\limits_{\{\lambda\}}\int\limits_{\{\alpha\}} F(y,\ \theta,\ \lambda)\, w_t(\lambda,\ \alpha \mid \theta)\, d\lambda\, d\alpha$$

The conditional density $w_t(\lambda,\ \alpha \mid \theta) = w_t(\lambda \mid \theta,\ \alpha)\, w_t(\alpha \mid \theta)$ is determined by the solution to the equations

$$\frac{\partial}{\partial t} w_t(\lambda \mid \theta,\ \alpha) = \mathbf{L}_\lambda w_t(\lambda \mid \theta,\ \alpha) + w_t(\lambda \mid \theta,\ \alpha)\,[F-\bar{F}] \tag{5.60}$$

$$\frac{\partial}{\partial t} w_t(\alpha \mid \theta) = w_t(\alpha \mid \theta)\,[\langle F\rangle(\alpha,\ \theta)-\bar{\bar{F}}] \tag{5.61}$$

where

$$F \equiv F(y,\ \theta,\ \lambda),\quad \langle F\rangle(\alpha,\ \theta)=\int\limits_{\{\lambda\}} F w_t(\lambda \mid \alpha,\ \theta)\, d\lambda$$

We obtain, by Eq. (5.61)

$$w_t(\alpha,\ \theta) = p_\alpha(\alpha) \exp\left\{\int_0^t [\langle F\rangle(\alpha,\ \theta)-\bar{F}]\, dt\right\} \tag{5.62}$$

Eqs. (5.58)-(5.62) define the reception algorithm. In the transition to quasilinear filtration algorithms, the algorithm structure exhibits manifestly the devices of sharping the synchroparameter estimates.

Adaptive reception problem. In cases where initial distribution of the parameter α cannot be specified, but as regards the information parameter the problem is formulated as a Bayessian one, the adaptive form of the Bayes procedure may be used [255]. In Eqs. (5.59) and (5.60) we substitute $w_t(\theta \mid \alpha)$ for $w_t(\theta)$ and in the solutions to these equations we replace α by the estimate of maximum likelihood $\hat{\alpha}_t$ which is to be found from the condition

$$\max_{\alpha} \Lambda_t(\alpha) = \Lambda_t(\hat{\alpha}_t) \tag{5.63}$$

where

$$\Lambda_t(\alpha) = \int_0^t \langle F\rangle(\alpha)\, dt$$

$$\langle F\rangle(\alpha)=\int\limits_{\{\theta\}}\int\limits_{\{\lambda\}} F(y,\ \theta,\ \lambda)\, w_t(\theta,\ \lambda \mid \alpha)\, d\theta\, d\lambda$$

The solution to the problem is then obtained from Eq. (5.58) by the substitution of $w_t(\theta \mid \hat{\alpha}_t)$ for $w_t(\theta)$.

Quasicoherent signal reception in binary FSK with clock frequency fluctuation. We now look at the problem of binary FSK signal reception with fluctuation of synchroparameters. The information sequence of signals $\{S(i,\ \theta_k^{(i)},\ \tau_k,\ \lambda),\ i=1,\ 2;\ k=1,\ 2,\ \ldots\}$ is

received in the presence of the background of white noise with spectral density N. The information parameter θ_i takes on two values $\{\theta_k^{(1)}, \theta_k^{(2)}\}$ with the distribution $\{p_1, p_2\}$. The sequence τ_k is a normal Markovian sequence with transient density determined by the parameters $(R\tau_{k-1};\ (1-R^2)\,\sigma^2)$ and initial density determined by the parameters $(0,\ \sigma_\tau^2)$, where $\sigma_\tau \ll T$. The clock intervals are $t_k = kT + \tau_k$, where T is the average duration of an elementary pulse. The Markov process $\lambda = \lambda(t)$ used as a model of fluctuation of signal phase (frequency) is given by the equations

$$\dot{\varphi} = \Omega \tag{5.64}$$

$$\dot{\Omega} = -\alpha\Omega + n_\Omega(t) \tag{5.65}$$

where $\mathrm{E}\,\{n_\Omega(t)\, n_\Omega(t+\tau)\} = N_\Omega \delta(\tau)$

The representation of λ in the form of Eqs. (5.64) and (5.65) corresponds to normal signal frequency fluctuations with the equivalent spectral band α.

Now we go on to find the signal identification algorithm with a minimal error probability. For a fixed clock interval $[t_k, t_{k+1}]$ the logarithm of likelihood ratio functional is given by the relationship [261]

$$z_k = z_k(t_k,\ t_{k+1}) = \int_{t_k}^{t_{k+1}} y(t)\,(\bar{s}_1 - \bar{s}_2)\,dt \tag{5.66}$$

where

$$\bar{s}_i = \mathrm{E}\,\{s(t,\ \theta_k^{(i)},\ \tau_k,\ \lambda) \mid y_{t_k}^t,\ \theta_k^{(i)},\ \tau_k\}$$

With high posterior accuracy of parameter τ_k we represent z_k at points $t_k^0 = kT + \tau_k^0$, $t_{k+1}^0 = (k+1)\,T + \tau_k^0 R$ by the first term of the Taylor series

$$z_k = z_k(t_k^0,\ t_{k+1}^0) = \frac{2}{N}\int_{t_k^0}^{t_{k+1}^0} y(t)\,(\bar{s}_1 - \bar{s}_2)\,dt \tag{5.67}$$

where t_k^0, t_{k+1}^0 are the traced and extrapolated clock moments, respectively. To shape z_k according to Eq. (5.67), a knowledge is required of estimates of τ_k^0 and posterior distribution of parameter λ, or estimates of $\hat{\lambda}$ in case where the approximation $\bar{s}_i \approx s(t,\ \theta_k^{(i)},\ \tau_k^0,\ \hat{\lambda})$ is valid.

Consider the case of clock synchronization. The recurrent equations of estimates τ_k^0 are an outcome of the general theory of filtration of sequences [261]

$$\hat{\tau}_k = \tau_k^0 + K_\tau \frac{\partial}{\partial\tau}\Pi_k(\tau_k^0),\quad \tau_k^0 = R\hat{\tau}_{k-1} \tag{5.68}$$

For the problem at hand

$$\Pi_k(\tau_k^0) = \ln\,[0.25e^{z_1+z_2}\,\mathrm{ch}\,(z_1 - z_2)] \tag{5.69}$$

It follows from Eq. (5.69)

$$\frac{\partial}{\partial\tau}\Pi_k(\tau_k^0)=\sum_{i=1}^{2}\pi_i(\tau_k^0)\frac{d}{d\tau}z_i \tag{5.70}$$

where

$$z_i \triangleq z_i(\tau_k^0)=\frac{2}{N}\int_{t_k^0}^{t_{k+1}^0} y(t)\bar{s}_i\,dt, \quad i=1,\ 2$$

$$\pi_i(\tau_k^0)=p_i\exp z_i\Big/\Big(\sum_{j=1}^{2}p_j\exp z_j\Big) \tag{5.71}$$

is the posterior probability of states $Q_k^{(i)}$

$$K_\tau=\frac{(1-R^2)(1-\sigma_\tau^2\overline{\Pi}_k'')}{2R^2\overline{\Pi}_k''}\times\left[\left(1-\frac{4R^2\sigma_\tau^2\overline{\Pi}_k''}{(1-R^2)(1-\sigma_\tau^2\overline{\Pi}_k'')}\right)^{1/2}-1\right] \tag{5.72}$$

is the variance of estimate of τ_k (stationary).

It is easily shown that for a large signal-to-noise ratio

$$\overline{\Pi}_k''=\sum_{i=1}^{2}\pi_i(\tau_k^0)\frac{d^2}{d\tau^2}z_i\approx q\langle\omega^2\rangle \tag{5.73}$$

is the averaged (in observation noise) second derivative in τ_k at the point τ_k^0 where

$$\langle\omega^2\rangle=\int\omega^2\,|f(\omega)|^2\,d\omega\Big/\int|f(\omega)|^2\,d\omega$$

$f(\omega)$ is the spectral density of signal; q is the signal-to-noise ratio.

A distinctive feature of the algorithm is its dependence on $\pi_i = \pi_i(\tau_k^0)$, i.e. the availability of solution feedback. This accounts for the difference of an appropriate block diagram from the clock synchronization arrangement obtained using the maximum likelihood technique [486].

Consider now the case of carrier synchronization. Under conditions of high posterior accuracy the probability density for the parameter λ in its kth frame is given by

$$w_t^{(k)}(\lambda\mid\theta_k^{(i)},\ \tau_k)\approx w_t^{(k)}(\lambda\mid\theta_k^{(i)},\ \tau_k^0)$$

Using this expression, we write the differential equations for recurrent estimates of the parameter $\lambda=(\varphi,\ \Omega)$

$$\left.\begin{aligned}\frac{d}{dt}\hat{\varphi}_k&=\hat{\Omega}_k+K_{\varphi\varphi}\frac{\partial}{\partial\varphi}F(t,\ \hat{\varphi}_k,\ \tau_k^0)\\ \frac{d}{dt}\hat{\Omega}_k&=-\alpha\hat{\Omega}_k+K_{\varphi\Omega}\frac{\partial}{\partial\varphi}F(t,\ \hat{\varphi}_k,\ \tau_k^0)\end{aligned}\right\} \tag{5.74}$$

where

$$F(t,\ \hat{\varphi}_k,\ \tau_k^0)=[1/(1+\Lambda)]\,F_1+[\Lambda/(1+\Lambda)]\,F_2$$

$$F_i=(1/N)\,y(t)\,s(t,\ \theta_k^{(i)},\ \tau_k^0,\ \hat{\varphi}_k)$$

$$\Lambda=\Lambda(t)=\exp\frac{1}{N}\int_{t_k^0}^{t} y(u)\,(\bar{s}_2-\bar{s}_1)\,du$$

$$=\exp\frac{1}{N}\int_{\tau_k^0}^{t}(F_2-F_1)\,du$$

$K_{\varphi\varphi}$, $K_{\varphi\Omega}$ are the elements of second moment matrix for the distribution of estimates λ which under stationary measurement conditions obey the set of equations

$$K_{\varphi\varphi}^2F_{\alpha\varphi}+2K_{\varphi\Omega}=0$$

$$K_{\varphi\varphi}^2F_{\varphi\varphi}-2\alpha K_{\Omega\Omega}+N_\Omega/2=0$$

$$K_{\varphi\Omega}K_{\varphi\varphi}F_{\varphi\varphi}-\alpha K_{\varphi\Omega}+K_{\Omega\Omega}=0$$

where

$$F_{\varphi\varphi}=\frac{\partial^2}{\partial\varphi^2}F(t,\ \hat{\varphi}_k,\ \tau_k^0)$$

The reception algorithm is defined by Eqs. (5.67), (5.68) and (5.74).

Adaptive clock synchronization. Let a sequence $\{s_k\}$ of signal in M-ary FSK with elementary symbol duration T_k be given in the presence of additive white noise with spectral density N: $s_k = s_k(t - t_k,\ \theta_k,\ \mu_k)$, $k = 1, 2, \ldots$; $t \in [t_k,\ t_{k+1}]$. Here θ_k is an information parameter determining the frequency shift keying of the signal, $\theta_k = \theta_k^{(1)}, \theta_k^{(2)}, \ldots, \theta_k^{(M)}$ with a distribution of $p_\theta(x)$; μ_k are immaterial random parameters $[\mu_k = (A_k,\ \varphi_k)]$, A_k, φ_k are the amplitude and phase of elementary signal. The sequence of moments $\{t_k\}$ will be given by the relationship $t_{k+1} = t_k + T_{k+1}$, where T_{k+1} obeys the recurrent equation

$$T_{k+1}=T_k+\alpha+\tau_k$$

$$\tau_{k+1}=R\tau_k+\sigma\sqrt{1-R^2}\,\xi_k \tag{5.75}$$

where α is a random parameter constant in the course of a communication session. In receiving discrete information signals, the dependence of them on the synchroparameter responsible for timing (elementary signal onset times).

The variation of the quantity T_k in the communication session is specified to a random parameter α. The sequence $\{\tau_k\}$ is controlled by the shaping instability at the sending end, irregularities of the medium of the signal propagation channel, and so on. Usually the assumption is valid that an approximation of a real sequence with a normal Markovian one is possible.

As is indicated above, the reception algorithm is associated with the formation of posterior density of distribution of a signal synchroparameter. Considering here only this part of the algorithm, we will obtain an approximate algorithm for filtering quasilinearly the synchroparameter, this algorithm being realized in a clock synchronizer. Thus, under the conditions discussed above the distinguishing of signals calls for the formation of posterior distribution of parameter (T_k, τ_k, α_k) from the observation of signals s_k, specified for $t \in [t_k, t_k + T_{k+1}]$, where

$$T_{k+1} = T_k + \alpha_k + \tau_k$$

$$\tau_{k+1} = R\tau_k + \sigma\sqrt{1-R^2}\,\xi_k$$

$$\alpha_{k+1} = \alpha_k$$

and ξ_k is the sequence of independent random quantities; $\mathrm{E}\,\{\xi_k\} = 0$, $\mathrm{E}\,\{\xi_k^2\} = 1$, R and σ^2 are the known quantities.

The initial conditions are set by specifying the distributions $p_\tau(x)$, $p_\alpha(x)$ for τ_0, α_0. The problem stated is a problem of estimating the Markovian parameter (T_k, τ_k, α_k) from the sequence of observations within the clock intervals $[t_k, t_{k+1}]$.

With fixed t_k, θ_k, and μ_k we write the logarithm of likelihood ratio functional for the parameter T_{k+1}

$$\Pi_k(T_{k+1}, \theta_k, \mu_k) = \frac{2}{N}\int_0^{T_{k+1}} y_{t-t_k} s_k(t-t_k, \theta_k, \mu_k)\,dt - \frac{1}{N}\int_0^{T_{k+1}} s_k^2(t-t_k, \theta_k, \mu_k)\,dt \tag{5.76}$$

Assuming $\mu = (A_k, \varphi_k)$, where A_k, φ_k are the amplitude and phase of radio pulse with distribution density

$$p_k(A_k, \varphi_k) = \frac{1}{2\pi}\frac{A}{\sigma_A}\exp(-A^2/2\sigma_A^2) \tag{5.77}$$

and designating the averaging over this distribution by

$$\Pi_k(T_{k+1}, i) \equiv \ln\langle\exp\Pi_k(T_{k+1}, \theta_k = i, \mu_k\rangle_{p_k(A_k, \varphi_k)}$$

we arrive at

$$\Pi_k(T_{k+1}, i) = -\ln(1+\rho(T_{k+1})) + \varepsilon_k^2(T_{k+1}, i)/2\,[1+\rho(T_{k+1})] \tag{5.78}$$

where

$$\rho(T_{k+1}) = \left\langle \frac{1}{N}\int_0^{T_{k+1}} s_k^2(t-t_k, \theta_k = i, \mu_k)\,dt \right\rangle_{p_k(A_k, \varphi_k)}$$

is the signal-to-noise ratio for one signal element (it is only natural to assume its independence of the value of $\theta_k = i$); $\varepsilon_k^2 (T_{k+1}, i) = X_k^2 (T_{k+1}, i) + Y_k^2 (T_{k+1}, i)$, $X_k(\cdot, \cdot)$, $Y_k(\cdot, \cdot)$ are quadrature components of correlation integrals. Further we have

$$\Pi_k (T_{k+1}) = \ln \langle \exp [\Pi_k (T_{k+1}, i)] \rangle_{p_\theta(x)}$$

$$- \ln \left\{ \sum_{i=1}^{M} p_i \exp [\Pi_k (T_{k+1}, i)] \right\} = \ln \left\{ \frac{1}{1+\rho(T_{k+1})} \sum_{i=1}^{M} p_i \right.$$

$$\left. \times \exp [\varepsilon_k^2 (T_{k+1}, i)/2 (1+\rho (T_{k+1}))] \right\} \tag{5.79}$$

where

$$p_\theta (x) = \{p_i, \ i = 1, 2, \ldots, M\}$$

The quasilinear synchronization algorithm is described by the set of equations [261]

$$\left.\begin{aligned} \hat{T}_{k+1} &= T_{k+1}^e + \overline{K}_{TT}(k) \Pi_k' (T_{k+1}^e) \\ \hat{\tau}_{k+1} &= \tau_{k+1}^e + \overline{K}_{T\tau}(k) \Pi_k' (T_{k+1}^e) \\ \hat{\alpha}_{k+1} &= \alpha_{k+1}^e + \overline{K}_{T\alpha}(k) \Pi_k' (T_{k+1}^e) \end{aligned}\right\} \tag{5.80}$$

where

$$T_{k+1}^e = \hat{T}_k + \hat{\alpha}_k + R\hat{\tau}_k, \quad \tau_k^e = R\hat{\tau}_k, \quad \tau_k^e = R\hat{\tau}_k$$

$$\alpha_{k+1}^e = \hat{\alpha}_k, \quad \Pi_k' (T_{k+1}^e) = \frac{d}{dT} \Pi_k (T) \big|_{T=T_{k+1}^e}$$

$\overline{K}_{TT}(k)$, $\overline{K}_{T}\ (k)$, $\overline{K}_{T\alpha}(k)$ are the elements of estimate covariation matrix $\overline{K}(k)$ satisfying the recurrent equations

$$\overline{K}^{-1}(k) = [D + \overline{AK}(k-1) A^T]^{-1} + \overline{\Pi}'' \tag{5.81}$$

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\tau^2 (1-R^2) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & R & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \overline{\Pi}_k'' = \begin{pmatrix} \overline{\Pi}_k' & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

It follows from Eq. (5.79)

$$\Pi_k' (T_{k+1}^e) \sum_{i=1}^{M} \pi_i \frac{d}{dt} \Pi_k (T_{k+1}^e, i) \tag{5.82}$$

where

$$\pi_i = p_i \exp [\Pi_k (T^e_{k+1},\ i)] \Big/ \sum_{j=1}^{M} p_j \exp [\Pi_k T^e_{k+1},\ j]$$

is the posterior probability of the state $\theta_k = i$.

Eq. (5.82) describes the discriminator configuration in the synchronization algorithm of Eq. (5.80).

5.8. Synthesis of Signal Discrimination Algorithms Based on Poly-Gaussian Models

Fundamentals of method. A random event is called poly-Gaussian if the pertinent probability measure and density (if any) can be expressed in terms of mixtures of normal (Gaussian) quantities. As poly-Gaussian models are in no way connected with the superposition principle in sample space, they are applicable both in linear and non-linear problems. So, the poly-Gaussian representation of a narrowband process results in a poly-Rician representation by an envelope. In particular, for a quasideterminate signal with a random amplitude a and phase φ

$$s(t) = a \cos (\omega t + \varphi) = u \cos \omega t + v \sin \omega t \quad \text{for} \quad a \geqslant 0 \tag{5.83}$$

with an arbitrary distribution $w(a, \varphi)$ from poly-Gaussian representations of its quadrature components $u = a \cos \varphi$ and $v = a \sin \varphi$, there follows a poly-Rician representation a [286]

$$w(a) = \sum_k q_k \frac{a}{\sigma_k^2} \exp \left\{ -\frac{a^2 - \sigma_k^2}{2\sigma_k^2} \right\} I_0 \left(\frac{\alpha_k a}{\sigma_k^2} \right),$$
$$\sum_k q_k = 1 \tag{5.84}$$

A practical utilization of poly-Gaussian models in synthesizing the reception algorithms requires the final mixture to be identifiable, which is known [450] to be always the case with normal components. The representation of a wide class of distributions by mixtures of normal distributions is most common in a number of problems in communication theory, pattern recognition theory, applied statistics [137, 181, 450, 438].

Now we take a problem of recognition of signals $s_k(t)$, $0 \leqslant t \leqslant T$, $k = 1, \ldots, m$ with prior probabilities p_k, $\sum_{k=1}^{M} p_k = 1$ and arbitrary multidimensional densities $w^{(k)}{}_N (x_1, \ldots, x_N)$ for an additive interference $\xi(t)$ with arbitrary multidimensional density

$w_N^{(0)}(x_1, \ldots, x_N)$. Replacing the densities $s_k(t)$ and $\xi(t)$ by mixtures of normal densities $w_N^{(k_\nu)}(x_1, \ldots, x_N)$

$$w_N^{(k)}(x_1, \ldots, x_N) \approx \sum_{\nu=1}^{n_k} q_\nu^{(k)} w_N^{(k\nu)} \quad (x_1, \ldots, x_N),$$
$$k=0, 1, \ldots, m \tag{5.85}$$

gives that the sum $x(t) = s_k(t) + \xi(t)$ is a poly-Gaussian process with the density

$$w_N^{(0k)}(x_1, \ldots, x_N) \approx \sum_{\nu=1}^{n_{0k}} q_\nu^{0k} w_N^{(0k\nu)}(x_1, \ldots, x_N),$$
$$k=1, \ldots, m$$

where $n_{0k} = n_0 n_k$, $(q^{0k)} = q_i^{(0)} q_j^{(k)}$. Here the numeration order for the resulting process components is unessential, and values of index $\nu = 1, \ldots, n_{0k}$ can, in particular, be distributed by numbering in succession the elements of the rectangular $n_0 \times n_k$ matrix, the line numbers being $i = 1, \ldots, n_0$, and column numbers $j = 1, \ldots, n_k$.

The total probability of detection error will be minimal if the region X_k of decision making as to the presence of signal $s_k(t)$ will be given by

$$p_k \sum_{\nu=1}^{n_{0k}} q_\nu^{(0k)} l_\nu^{(k)}[x(t)] = \max_{1 \leqslant i \leqslant m} p_i \sum_{i=1}^{n_{0i}} q_\nu^{(0i)} l_\nu^{(i)}[x(t)],$$
$$k=1, \ldots, m \tag{5.86}$$

where

$$l_\nu^{(k)}[x(t)] = \lim_{N\to\infty} w_N^{(0k\nu)}$$
$$\times (x_1, \ldots, x_N)/w_N^{(b)} \quad (x_1, \ldots, x_N) \tag{5.87}$$

are the conditional likelihood functionals; $w_N^{(b)}(x_1, \ldots, x_N)$ is the auxiliary density such that for all k there exist limits of Eq. (5.87).

Optimal discrimination in the presence of random impulse noise. Consider the problem of discrimination of signals (important, for instance, for communications with moving objects) $s_k(t) = \xi_{s\,k} \operatorname{Re}[u_{s\,k}(t) \exp(2\pi ift + \varphi_{s\,k})]$, $t \in T$, $k = 1, \ldots, m$ in the presence of random pulsed interference (RPI) $u_{\text{int}}(t) = \xi_{\text{int}\,n} \operatorname{Re}[u_{\text{int}\,n} s_{\text{int}\,n}(t) \exp(2\pi ift + \varphi_{\text{int}\,n})]$, $t \in T_{\text{int}}$, $n = 1, \ldots, h$ with arbitrary predetermined probability densities $w(\xi_{s\,k})$ and $w(\xi_{\text{int}\,n})$ of amplitude fluctuations in the presence of noise with uniform (within the radio link frequency band) power spectral density $N_0 \in (N_0^{\min}, N_0^{\max})$ that varies at random with probability density $w^{(n)}(N_0)$ [284]. The probability of appearance of

signals is $p^{(sk)}$, $k = 0, \ldots, m$, and of RPI—$p^{(\mathrm{int}\, n)}$, $n = 0, \ldots, h$ (zero indices signify no signals and RPI). In situations of practical interest the RPI flow densities λ_n are such that $\max\limits_n \lambda_n (T_{\mathrm{int}} + T) \ll 1$ and the resultant RPI flow is ordinary, random initial phases φ_{sk} and $\varphi_{\mathrm{int}\, n}$ are equiprobable, and pulse amplitude fluctuations of signals and interference, as follows from [286], may be represented in terms of poly-Rician densities similar to Eq. (5.84) with parameters $q_{s\, ki}, \alpha_{s\, ki}, \sigma_{s\, ki}$, $i = 1, \ldots, n_{s\, k}$, $k = 1, \ldots, m$ and $q_{\mathrm{int}\, nl}, \alpha_{\mathrm{int}\, nl}, \sigma_{\mathrm{int}\, nl}$ $l = 1, \ldots, n_{\mathrm{int}\, n}$, $h = 1, \ldots, h$, respectively. Poly-Rician expressions are easily obtained for all the conditional functionals entering Eq. (5.86) [285]. So, for example, the functional of likelihood ratio under the hypothesis of the interaction of the signal $s_k(t)$ and a pulse belonging to the nth RPI flow will be given by

$$
\begin{aligned}
l_n^k [x(t),\ t_i] = {} & \frac{1}{T_{\mathrm{int}\, n} + T} \int\limits_{-T_{\mathrm{int}\, n}}^{T} \frac{N_0^* \, dt_i}{N_0^* - 2\sigma_{sk}^2 U_{sk}^2 Y_{kn}(t_i)} \\
& \times \exp\left\{ \frac{1}{N_0^*} \frac{2\sigma_{sk}^2 U_{sk}^2 Z_{kn}[x(t),\ t_i]}{N_0^* + 2\sigma_{sk}^2 U_{sk}^2 Y_{kn}(t_i)} \right\} \\
& \times \sum_{i=1}^{n_{sk}} q_{sk\, i} \sum_{l=1}^{n_{\mathrm{int}\, n}} q_{\mathrm{int}\, nl} \exp\left\{ - \frac{\alpha_{sk\, i}^2 U_{sk}^2 Y_{kn}(t_i)}{N_0^* + 2\sigma_{sk}^2 U_{sk}^2 Y_{kn}(t_i)} \right. \\
& \left. - \frac{\alpha_{\mathrm{int}\, nl}^2 U_{\mathrm{int}\, n}^2 Y^{kn}(t_i)}{N_0^* + 2\sigma_{\mathrm{int}\, n}^2 U_{\mathrm{int}\, n}^2 Y_{kn}(t_i)} \right\} I_0 \left\{ \frac{\alpha_{sk\, i} U_{sk} Z_{kn}[x(t),\ t_i]}{N_0^* - 2\sigma_{sk}^2 U_{sk}^2 Y_{kn}(t_i)} \right\} \\
& \times I_0 \left\{ \frac{\alpha_{\mathrm{int}\, nl} U_{\mathrm{int}\, n} Z^{(kn)}[x(t),\ t_i]}{N_0^* + 2\sigma_{\mathrm{int}\, n}^2 Y^{(kn)}(t_i)} \right\} \\
& \times I_0 \left\{ \frac{\alpha_{sk\, i} \alpha_{\mathrm{int}\, nl} U_{sk} Z_{kn}[x(t),\ t_i]}{N_0^* + 2\sigma_{sk}^2 U_{sk}^2 Y_{kn}(t_i)} \right\}
\end{aligned}
\tag{5.88}
$$

where

$$
Z_{kn}[x(t),\ t_i] = \frac{1}{2} \left| Z_{k0}[x(t)] - \frac{2\sigma_{\mathrm{int}\, n}^2 U_{\mathrm{int}\, n}^2 Y_{k0}(t_i)}{N_0^* + 2\sigma_{sk}^2 U_{sk}^2 Y_{k0}(t_i)} Z_{k0}[x(t),\ t_i] \right|
$$

$$
Z^{(kn)}[x(t),\ t_i] = \frac{1}{2} \left| Z_{k0}[x(t),\ t_i] - \frac{2\sigma_{sk}^2 U_{sk}^2 Y_{k0}(t_i)}{N_0^* + 2\sigma_{\mathrm{int}\, n}^2 U^2_{\mathrm{int}\, n} Y_{k0}(t_i)} Z_{k0}[x(t)] \right|
$$

$$
Y_{kn}(t_i) = \frac{1}{2} \left| Y_{k0} - \frac{2\sigma_{\mathrm{int}\, n}^2 Y_{k0}(t_i)}{N_0^* + 2\sigma_{sk}^2 U_{sk}^2 Y_{k0}(t_i)} Y_{k0}(t_i) \right|
$$

$$
Y^{(kn)}(t_i) = \frac{1}{2} \left| Y_{k0}(t_i) - \frac{2\sigma_{sk}^2 U_{sk}^2 Y_{k0}(t_i)}{N_0^* + 2\sigma_{\mathrm{int}\, n}^2 U_{\mathrm{int}\, n}^2 Y_{k0}(t_i)} Y_{k0} \right|
$$

$$Z_{k0}[x(t)] = \frac{1}{2}\left|\int_0^T X(t)\, S_{sk}(t)\, dt\right|; \; Y_{k0} = \frac{1}{2}\int_0^T |S_{sk}(t)|^2\, dt$$

$$Z_{k0}[x(t),\ t_i] = \frac{1}{2}\left|\int_0^T X(t)\, S_{sk}^{-}(t - t_i)\, dt\right|;$$

$$Y_{k0}(t_i) = \frac{1}{2}\int_0^T |S_{sk}(t - t_i)|^2\, dt$$

An estimate of the noise spectral density N_0^+ may be obtained by the conventional way once the hypothesis is accepted that signals are absent.

Thus, the optimal algorithm of detection—discrimination of signals with random power of noise and h ordinary RPI flows—is multicorrelative in the general case of arbitrary fluctuations of information-bearing and disturbing pulses [285]. It reduces to a selection of the largest among $m + 1$ weighted sums for the $n + 1$ results of single-type non-linear multichannel transformations of respective correlation integrals. These channels correspond to the Rician components of fluctuations of signals and interference, and the transformation parameters depend on the power of actual noise, its estimation being performed upon accepting the hypothesis that impulses are absent in the input process.

We note the invariance of configuration of the algorithm derived: in changing the pulse fluctuation probability density the algorithm structure remains invariable, with only parameters of individual channels varying. Block diagrams of poly-Gaussian receivers with the analysis of their noise immunity are provided in [284-286].

Adaptation of poly-Gaussian receivers. The arrangement of poly-Gaussian receivers is quite convenient as regards adaptation to the parameters of both signals and interference. The mean values, correlation functions and prior probabilities of components are just the parameters, the estimation of which (or some of them) would ensure the receiver adaptation.

So, for example, the task of discrimination of deterministic signals $s_k(t)$, $0 \leqslant t \leqslant T$, $k = 1, \ldots, m$ with probabilities $p^{(k)}$ and additive interference $\xi(t)$ with unknown prior distribution consists in achieving the estimates of probability q_ν^* based on the apparent realization of $x(t)$, $0 \leqslant t \leqslant T$, before the decision is made as to the actual signal.

The results of the decision about the signal are fed to the receiver output, and the results of decision about the interference component are directed to appropriate storage counters, whence the stored normalized quantities q_ν^* are transferred to the storage registers to be

used in the course of the next adaptation cycle as probabilities of availability of interference components in its poly-Gaussian representation.

Note that this algorithm shows one of the advantages of poly-Gaussian representations in communication theory, i.e. the identity of operations with video signals in all the channels both in signal estimation circuits, and in interference component estimation circuits. Thus, if in the discrimination problem considered above the average power is a priori unknown for not only the noise but also the pulsed component of total interference, this power being dictated by densities λ_n of partial RPI, then the desired estimates λ_n^* may also be obtained using the techniques common in estimation theory.

Block diagrams of adaptive poly-Gaussian receivers for practically important cases of discrimination of pulsed signals with RPI are available in references [284-286].

Chapter

6

Distortion and Crosstalk in Multichannel Radio Communication Systems

6.1. Introduction

Background. The development of earlier multichannel communication systems necessitated the fundamental research into distortion and crosstalk. Multichannel telephone systems based on the frequency division multiplexing and single-sideband transmission were developed and introduced on cable lines in the 1930s. The capacity of these systems increased rapidly ever since. In 1936, came 60 kHz, 12-channel communication systems operated on symmetrical cable links, and as early as in 1939 a 240-channel coaxial-cable telephone system with a frequency range up to 1 MHz was put into operation. As the capacity increased, intervals between repeaters decreased and the number of repeaters in the link grew. The problem of crosstalk, which arises due to amplifier nonlinearity, became increasingly more important. All the above called for a theorectical analysis of multichannel transmission, but at the time it could not be done because an adequate mathematical technique was not available. In theoretical studies of telecommunications conventional representation of signals as harmonic oscillations, i.e. deterministic processes were used, whereas a multichannel signal is essentially a random process.

The first attempt at a statistical approach to crosstalk analysis was undertaken in 1939 by Lubny-Gertsyk [183]. He represented a multichannel signal as a random variable whose probability distribution tends to normal, according to the central limit theorem, as the number of channels increases. Unfortunately, this work found no response with the communication people. After the fundamental investigations into the theory of random processes have been published the communication people were rather wary in accepting this theory. In 1945 Brockbank and Wass [329] analyzed the intermodulation noise in a multichannel system due to repeater nonlinearity. They represented multichannel signal by a sum of harmonic oscil-

lations at different frequencies and derived formulas which gained wide acceptance in practice.

The 1940s saw first multichannel radio relay systems. These systems used the same EDM technique and multiplexing equipment as in cable systems but for transmission, a radio signal frequency-modulated by the multichannel signal was used. The different way of transmission made the theoretical study of interference and distortions still more complicated than for wire communications, yet the development of radio relay systems and an increase in their capacity compel the analysis to be done.

To study crosstalk in multichannel radio communication systems S. V. Borodich [35] used random process correlation theory and the mathematical model of multichannel signal in the form of a normal stationary random process with a uniform spectrum. This approach proved rather fruitful to yield later a number of significant practical results [35].

Types of distortion and crosstalk in multichannel radio communication systems. Consider multichannel FDM-FM systems. These are employed in radio relay and satellite communications. Interference in telephone channels of these systems consists of two components: thermal noise and crosstalk. The crosstalk is caused by non-linear distortions of a multichannel signal in the link. In many cases crosstalk is a major limitation to the capacity of the link, and its control is an important problem.

At the very outset of the radio-relay communication it was widely believed that with FM modulation, non-linear distortions of the baseband signal and hence crosstalk could be kept to a minimum only through the linearization of modulator and demodulator characteristics and the group delay equalization in RF channel. As the radio-relay technology advanced and system capacity increased, ever new sources of crosstalk were coming to light. It appeared that non-linear distortion might occur in all the components of the link resulting in crosstalk in the channels. To reduce the interference, an adequate equalization of characteristics of link components and adequate selection of their parameters are needed.

In the baseband circuit handling a multichannel signal, crosstalk arises due to nonlinearity of transfer characteristics which show instantaneous output voltages or frequency deviations of radio signal as a function of instantaneous input voltages. This interference has been treated theoretically in [34, 329, 334].

In an RF channel where a radio signal frequency-modulated by a multichannel signal passes, the crosstalk is for the most part caused by the nonuniformity of the frequency response and the nonlinearity of the phase characteristics. In a simplified form this type of crosstalk was first considered assuming small distortions and approximating the RF channel characteristics by a polynomial in

powers of deviation from the centre frequency. A more general analysis is given in [458, 457]; however the examination was limited by the third order distortions even with this truncation the results obtained are too complicated to be employed in engineering calculations.

There are many other sources of intermodulation noise in an RF channel. A spurious amplitude modulation arising when a FM signal passes through a RF channel is converted into phase modulation in some elements of the circuit; the pattern of this modulation is distorted, thus resulting in crosstalk [349]. An incomplete suppression of amplitude modulation in the frequency detector limiter also causes crosstalk. In poorly matched antenna feeders, signals reflected from feeder ends and from section joints are responsible for echo-signals delayed in relation to the main signal. Coming to the frequency detector together with the main signal, the echo-signals give rise to nonlinear distortions in a multichannel signal. The echo-signals may arise on pathes between the antennas of adjacent stations due to the reflection of radio waves from the earth or troposhere irregularities. The nonlinear distortions due to the combined effect of a wanted FM signal and a delayed echo-signal on a frequency detector were first studied by V. A. Smirnov [247]. In [324] the correlation theory of random processes and a multichannel signal model in the form of a normal stationary random process are used to evaluate crosstalk caused by echo-signals in the antenna feeder. Unfortunately, estimation formulas have only been obtained for several special cases of no much interest for practice.

Still another source of crosstalk is radio interference which may by internal, i.e. originating within the multichannel system itself, and external, i.e. caused by other systems. Medhurst et al. [439] were the first to investigate crosstalk in a multichannel radio relay system generated by an interfering radio signal of a similar system, later the paper by Hamer [381] was published. Quantitative results in these papers have been obtained for several particular cases only. Ruthroff [465] and Chibi [293] considered another mechanism of radio interference effect when the interfering signal converts the amplitude modulation of the wanted signal into a phase modulation in the amplitude limiter of the receiver. The effect of the amplitude-modulated interference was the subject of the paper by Ju. F. Marchenko [185]. All these papers assumed the frequency detector of the receiver being ideal. L. Ja. Kantor et al. [122] studied how crosstalk was effected with the possible departure of frequency detector responses from ideal. It is for the first time that the nonlinear character of the frequency detector and inadequate suppression of the amplitude modulation in the receiver limiter have been identified and studied as the source of additional crosstalk due to interfering signal.

In satellite communication systems, in addition to the mentioned,

there is yet another cause for crosstalk—an interaction of several earth stations signals amplified by the common nonlinear satellite transponder. This effect is analyzed in a great number of works, one of the first papers being [1, 488, 502]. Monograph [35] summarizes the results of investigations into all the above phenomena.

6.2. Statistical Parameters of Signals and Crosstalk in Multichannel Radio Communication Systems

Power spectrum of a radio signal. A multichannel signal is the sum of frequency converted individual signals transmitted over channels and in many cases is a stationary random process. The fact that the process is stationary manifests itself during a rather long period of time which normally is the busy hour (BH) during which the number of active channels in the system is maximum. As the number of channels increases, the probability distribution of instantaneous values of a multichannel signal tends to normal. It has been shown experimentally that the statistical distribution may be regarded as normal if the number of voice channels is more than 60. A mathematical model of the multichannel signal is a normal stationary random process with zero mean and variance σ_U^2. The power spectrum of the process is thought to be uniform and bounded within a frequency band of $\Omega_1 \leqslant \Omega \leqslant \Omega_2$. The correlation function of the process, $K_U(\tau)$, is then determined from the Wiener-Khinchin theorem. Also, the process is characterized by the average power at BH, P_{av}, or average power level expressed in decibels.

The multichannel signal $U(t)$ modulates the frequency of the radio signal, and the phase shift of radio signal is given by the integral with a variable upper limit

$$s(t) = C_m \int_{-T}^{t} U(\xi)\, d\xi \tag{6.1}$$

where C_m is the slope of modulator response; $-T$ is the instant of modulation switch-on. The process $s(t)$ is also normally distributed, but it is nonstationary, as has been noted by B. R. Levin [162]. If, however, the modulation is on for sufficient time, T may be supposed to lie somewhere at infinity, the process becomes stationary in a wide sense.

A radio signal frequency-modulated by a multichannel signal

$$v(t) = u_0 \cos[\omega_0 t + s(t)] \tag{6.2}$$

is a nonstationary random process. Its time-averaged correlation function [162] is

$$\langle K_v(\tau)\rangle = \frac{u_0^2}{2} \exp - [K_s(0) - K_s(\tau)] \cos \omega_0 \tau \tag{6.3}$$

and the mean power spectrum is obtained from the equation [162]

$$\langle F(\omega)\rangle = \frac{u_0^2}{2} \exp - [K_s(0) - K_s(\infty)]$$
$$\times \left\{ \delta(\omega - \omega_0) + \frac{1}{\pi} \int_0^\infty [\exp - [K_s(\tau) - K_s(\infty)] - 1] \right.$$
$$\left. \times \cos(\omega - \omega_0)\tau \, d\tau \right\} \tag{6.4}$$

The spectrum has a discrete component at frequency ω_0 the power of which is dependent on the bound frequency ratio $\beta = \Omega_2/\Omega_1$ of the multichannel signal spectrum, and on the rms modulation index squared

$$M_{\mathrm{rms}}^2 = C_m K_U(0)/\Omega_2^2 = (\Delta\omega_{\mathrm{rms}}/\Omega_2)^2 \tag{6.5}$$

If $M_{\mathrm{rms}}^2 \beta \gg 1$, then the discrete component in the spectrum vanishes. With a large rms modulation index, i.e. $M_{\mathrm{rms}}^2 \gg 1$, the power spectrum of radio signal is independent of the spectrum shape of multichannel signal and has the form [162]

$$\langle F_{FM}(\omega)\rangle \approx \left(1/\Delta\omega_{\mathrm{rms}}\sqrt{2\pi}\right) \exp\{-0.5\,[(\omega - \omega_0)/\Delta\omega_{\mathrm{rms}}]^2\} \tag{6.6}$$

To calculate the spectrum of an FM signal with small modulation index, the integral is estimated asymptotically by the method of steepest descents (which is convenient for large values of β), or else convolution integrals may be taken numerically. Figure 6.1 shows by way of example the power spectrum of a radio signal in a radio-relay system with 1920 telephone channels at $M_{\mathrm{rms}}^2 = 0.01625$.

Crosstalk in telephone channel. In the presence of nonlinear distortions of a baseband signal, there appear products of nonlinearity, and a part of them gets into a band occupied by the baseband signal and produces crosstalk. To determine the crosstalk value, the spectral analysis is most practicable to ascertain the distribution of nonlinearity products within the bandwidth of a multichannel signal. At the baseband circuit output the multichannel signal is presented as a sum of undistorted process and products of nonlinear distortions

$$U^*(t) = U(t) + \varepsilon(t) \tag{6.7}$$

At first the autocorrelation function $\varepsilon(t)$ of the process is found and then the part of it which corresponds to the products coherent with $U(t)$, is excluded as it does not give rise to crosstalk. This part is a linear function of $K_U(\tau)$. Further, the power spectrum of nonlinearity products, $F_\varepsilon(\Omega)$, and the crosstalk power in the channel with centre frequency Ω_{ch} and bandwidth $\Delta\Omega_{ch}$ are found. The bandwidth of a telephone channel is much narrower than the bandwidth of the baseband circuit, therefore the intermodulation noise

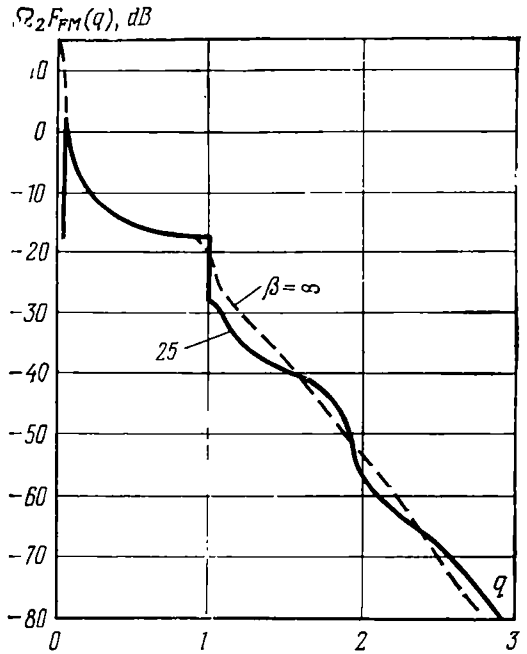

Fig. 6.1. Power spectrum of radio signal

spectral density may be thought constant within the bandwidth of one channel. The psophometrically weighted power of crosstalk in a telephone channel with loading resistance R is then

$$P_{nl} = F_{\varepsilon}(\Omega_{ch})\,\Delta\Omega_{ch} k_p^2 / R B^2(\Omega_{ch}) \tag{6.8}$$

where k_p is the psophometric weighting coefficient, $B(\Omega)$ is the transfer factor of a preemphasis network.

Equation (6.8) is conveniently transformed so as to express crosstalk power in the telephone channel in picowatts as the zero relative level

$$P_{nl} = 10^9 k_p^2 \,[\Delta F_{ch}/\Delta F B^2(\Omega_{ch})]\,\mathscr{F}_g(\Omega_{ch}) \tag{6.9}$$

where $\Delta F = (1/2\pi)(\Omega_2 - \Omega_1)$ is the baseband signal bandwidth, $\mathscr{F}_g(\Omega)$ is a dimensionless function proportional to the intermodulation noise power spectral density. The function $\mathscr{F}_g(\Omega)$ is derived for the transfer function of the baseband circuit represented by a polynomial of the fifth order. In many practical cases the third-order polynomial would suffice. In that case the relationship for $\mathscr{F}_g(\Omega)$ is very compact

$$\mathscr{F}_g(\Omega) = 4 \times 10^{0.2p_{av}} \times 10^{-0.1b_{2k}} y_2(\Omega) + 24 \times 10^{0.3p_{av}} \times 10^{-0.1b_{3k}} y_3(\Omega) \tag{6.10}$$

where p_{av} is the level of baseband circuit power in decibels; b_{2k}, b_{3k} are the attenuations of channel nonlinear components at the 2nd and 3d harmonics, in decibels, measured when the level of test signal at the channel input equals to the channel measurement level; $y_2(\Omega)$, $y_3(\Omega)$ are dimensionless functions in graphical representation.

If the baseband circuit is represented by an ideal limiter with a linear portion, then

$$\mathcal{F}_g(\Omega) = 10^{0.1p_{av}} \frac{2}{\pi} e^{-(U_0/\sigma_U)^2} \times \sum_{n=1}^{\infty} \frac{H_{2n-1}^2 (U_0/\sigma_U)}{(2n+1)!} y_{2n+1}(\Omega) \quad (6.11)$$

where U_0 is the limiting threshold, $H_{2n-1}(U_0/\sigma_U)$ is the Hermite polynomial.

Equation (6.11) describes the effect of the limitation of the multichannel signal peak values on crosstalk. The power of intermodulation noise is found to be insignificant at $U_0/\sigma_U\sqrt{2} > 3$, hence in calculating the crosstalk the transfer characteristic of the circuit might be well approximated within the range $-3\sigma_U\sqrt{2} \leqslant U(t) \leqslant 3\sigma_U\sqrt{2}$.

Crosstalk in RF IF circuit. When dealing with crosstalk occurring in RF IF circuits, it is convenient to represent the signal phase at the circuit output as the sum of the undistorted process and the phase error $s(t) + \theta(t)$ and seek further the correlation function of the phase error and its power spectrum. To derive the power spectrum of nonlinearity products at the demodulator output, it is convenient to use the theorem of spectrum of a derived process. Then the crosstalk power is

$$P_{nl} = 10^9 k_p^2 [\Delta F_{ch}/\Delta F B^2(\Omega_{ch})] (F_{ch}/\Delta f_{ch})^2 \mathcal{F}_B(\Omega_{ch}) \quad (6.12)$$

where Δf_{ch} is the rms deviation of radio signal frequency corresponding to a measurement tone level in the channel, $\mathcal{F}_B(\Omega)$ is a dimensionless function proportional to the spectral density of the phase error.

In all the cases considered, the expressions for the function $\mathcal{F}_B(\Omega)$ are rather complicated, so, for engineering applications, tables and curves of this function are calculated.

The analysis of crosstalk due to echo-signals in antenna feeders of a radio-relay link gives the following expression for the function $\mathcal{F}_B(\Omega)$:

$$\mathcal{F}_B(\Omega) = \sum_{i=1}^{n} \sum_{k=1}^{n} \{X_i X_k [G_{ik}(\Omega) - H_{ik}(\Omega)] + Y_i Y_k [G_{ik}(\Omega) + H_{ik}(\Omega)]\} \cos \Omega \frac{\tau_i - \tau_k}{2} \quad (6.13)$$

Here

$$X_i = \sum_{j=1}^{N_i} K_j \cos(\omega_0 t_j - \varphi_j),$$

$$Y_i = \sum_{j=1}^{N_i} K_j \sin(\omega_0 t_j - \varphi_j)$$

and K_j is the echo-signal amplitude, N_i is the number of echo-signals with delay time τ_i. The functions $G_{ik}(\Omega)$ and $H_{ik}(\Omega)$ are given by the integrals

$$G_{ik}(\Omega) = \frac{1}{\pi} \exp - [y(x_i) + y(x_k)]$$
$$\times \int_0^\infty \{\exp[-M_{\mathrm{rms}}^2 \eta(xx_ix_k)] - 1$$
$$+ M_{\mathrm{rms}}^2 \eta(x, x_i, x_k)\} \cos bx\,dx \tag{6.14}$$

$$H_{ik}(\Omega) = \frac{1}{\pi} \exp - [y(x_i) + y(x_k)]$$
$$\times \int_0^\infty \{\exp[M_{\mathrm{rms}}^2 \eta(xx_ix_k)] - 1$$
$$- M_{\mathrm{rms}}^2 \eta(x, x_i, x_k)\} \times \cos bx\,dx \tag{6.15}$$

$$x_i = \Omega_2 \tau_i, \quad x_k = \Omega_2 \tau_k, \quad x = \Omega_2 \tau, \quad b = \Omega/\Omega_2$$

$$\eta(x, x_i, x_k) = y(x - \mu_1)$$
$$+ y(x + \mu_1) - y(x - \mu_2) - y(x + \mu_2)$$

$$\mu_1 = (x_i - x_k)/2,$$

$$\mu_2 = (x_i + x_k)/2$$

$$y(x) = 0.4[x \operatorname{Si} x + \cos x - 1] - 1.35(\sin x/x)$$
$$+ 0.75 \frac{d^2}{dx^2}(\sin x/x) + 1.6$$

These functions were derived by numerical integration, and in [35] their plots on the plane x_i, x_k have been constructed for several values of $b = \Omega/\Omega_2$ and several values of M_{rms} most common in practice.

Transformation of wanted and interfering signals in video channels. Considering the wanted FM signal and an interfering signal as applied to an ideal frequency demodulator, the channel crosstalk has been shown to depend on an overlap of their spectra, the function

$\mathcal{F}_B(\Omega)$ being proportional to the convolution of these spectra

$$\mathcal{F}_B(\Omega)=\frac{1}{2}\frac{P_{int}}{P_s}\left\{\int_{-\infty}^{\infty}K^2(\omega_{int}-\Omega-u)\,F_s(\delta\omega-\Omega-u)\right.$$
$$\times K^2(\omega_{int}-u)\,F_{int}(u)\,du$$
$$+\int_{-\infty}^{\infty}K^2(\omega_{int}+\Omega-u)$$
$$\left.\times F_s(\delta\omega+\Omega-u)\,K^2(\omega_{int}-u)\,F_{int}(u)\,du\right\} \tag{6.16}$$

where P_s, P_{int} are the powers of the wanted and interfering signals at the receiver input; $F_s(\omega)$, $F_{int}(\omega)$ are their power spectra; $K(\omega)$ is the modulus of the transfer function from the receiver input to the frequency demodulator; $\delta\omega = \omega_{int} - \omega_s$ is the difference between carrier frequencies of the wanted and interfering signals. To calculate $\mathcal{F}_B(\Omega)$ from Eq. (6.16) the recourse is had to numerical integration.

For a real frequency demodulator, account must also be taken of the AM/FM conversion, incomplete suppression of AM by the limiter, and nonlinearity of the discriminator. If an interfering signal is amplitude-modulated (e.g. if the carrier frequency of a FM interfering signal lies on a slope of the receiver filter frequency response), then the above reasons are responsible for the appearance of crosstalk components independent of signals spectra overlapping and proportional to their squared power ratio.

Interferences which are due to amplification of several FM signals in a common nonlinear transponder are studied in much the same way. The computational formula for the transponder nonlinearity is given in terms of the ratio of a single signal power to the total power of the third-order nonlinearity products at the frequency of the signal. This ratio is obtained experimentally for four signals with subsequent interpolation on a specified number of signals.

To analyze the crosstalk caused by the network frequency response nonuniformity and phase response nonlinearity, these characteristics should be expressed analytically. They are usually approximated by the functions convenient for the analysis, which is only possible within a limited frequency band. The question as to how wide this band should be has been open for a long time. The solution has been found in the correlation theory whereby the required bandwidth within which the RF channel characteristics have influence on the crosstalk value has been determined.

The phase error introduced into the FM signal by a RF channel is given, in the general form, by

$$\theta(t)=\operatorname{arctg}\left[x_s(t)/X_c(t)\right] \tag{6.17}$$

$$x_s(t)=\int_{-\tau_b}^{\infty} H_c(\tau+\tau_b)\sin[s(t-\tau)-s(t)]\,d\tau$$

$$-\int_{-\tau_b}^{\infty} H_s(\tau+\tau_b)\cos[s(t-\tau)-s(t)]\,d\tau \tag{6.18}$$

$$X_s(t)=\int_{-\tau_b}^{\infty} H_c(\tau+\tau_b)\cos[s(t-\tau)-s(t)]\,d\tau$$

$$+\int_{-\tau_b}^{\infty} H_s(\tau+\tau_b)\sin[s(t-\tau)-s(t)]\,d\tau \tag{6.19}$$

where $H_c(\tau)$, $H_s(\tau)$ are the envelopes of impulse response of the RF channel.

From these relationships the known quasistationary approximation is readily obtained. If the impulse response attenuation is sufficiently fast, and is practically zero at $\tau > \tau_0$, and if at $\tau \leqslant \tau_0$ the approximate equality $s(t-\tau)-s(t) \approx -\tau s'(t) = -\tau C_m U(t)$ is valid, then from Eqs. (6.17) through (6.19) the quasistationary approximation results. It is applied under the conditions that within the necessary bandwidth defined above the power spectrum of the signal has the form of a Gaussian curve, and the channel characteristic has no singularities. These conditions are satisfied in low-capacity radio-relay systems and in satellite communication systems. In approximating the characteristics of the group delay time (GDT) by a second-degree polynomial in powers of detuning $\tau_b(F) = -\gamma_1 F + \gamma_2 F^2$, the quasistationary approximation gives a simple relationship

$$\mathcal{F}_B(\Omega)=4\pi^2\,[(1/2)\,\gamma_1^2\Delta f_{\mathrm{rms}}^4 y_2(\Omega)$$

$$+(2/3)\,\gamma_2^2\Delta f_{\mathrm{rms}}^6 y_3(\Omega)] \tag{6.20}$$

For high capacity radio-relay systems the quasistationary approximation cannot be used for the crosstalk analysis. For that case, two methods of calculation have been devised, both assuming small distortions. In addition, the first method is based on the assumption that the network impulse response has finite duration and that the rms modulation index is small, whereas the second one uses a well-known method of paired echoes. The frequency response of a network within the required bandwidth determined above is approximated by a Fourier series with a finite number of terms. In so doing, the signal at the output of a network is expressed as a sum of the input signal and paired echo-signals both leading and lagging the basic signal by a time proportional to the harmonic number in the Fourier series and having amplitudes proportional to the coefficients of that series. The second method gives the following rela-

tionship:

$$\mathcal{F}_B(\Omega) = [\sum_{n=1}^{N-1} B_n K_{odd}(n\Delta\omega_{pt}/2N)]^2$$
$$+ [\Delta f_{pt} \sum_{n=1}^{N-1} C_n \tau_{odd}(n\Delta\omega_{pt}/2N)]^2$$
$$+ [\sum_{n=0}^{N} A_n K_{ev}(n\Delta\omega_{pt}/2N)]^2$$
$$+ [\Delta f_{pt} \sum_{n=0}^{N} D_n \tau_{ev}(n\Delta\omega_{pt}/2N)]^2 \qquad (6.21)$$

Equation (6.21) includes even, K_{ev}, τ_{ev}, and odd, K_{odd}, τ_{odd}, channel responses samples (AFR and GDT) at discrete equidistant frequencies within the bandwidth $\Delta\omega_{pt} = 2\pi\Delta f_{pt}$. The coefficients A_n, B_n, C_n, and D_n (as products) are determined using the functions $G_{ik}(\Omega)$ and $H_{ik}(\Omega)$ [see Eqs. (6.14) and (6.15)]. If the tables of these coefficients are available, Eq. (6.21) is very convenient for calculations as it includes the samples of network characteristics at discrete frequencies only. Finding of the required number of samples, N, however involves some uncertainty.

Probability distribution of crosstalk power. In designing a multichannel communication link characteristics of all its components should be determined on the bases of a specified value of channel crosstalk. However, when the link is established the measured crosstalk may differ from the calculated value due to the random deviation in the RF channel characteristics. In various similarly designed links the measured powers of crosstalk will not be equal, therefore the crosstalk power at the link end should be regarded as a random variable assuming different values with different link realizations. Such an approach enables tolerable limits to be defined for the random dispersion of channel characteristics, using the distribution of that random variable.

It has been possible to use this approach for evaluating the power of crosstalk resulting from echo-signals in antenna feeders of radio-relay link and from a random dispersion in RF channel characteristics. The function $F_B(\Omega)$ defined by Eq. (6.13), should be considered as a random variable

$$\mathcal{F}_B(\Omega) = \rho_{2n} = \mu_n + \nu_n \qquad (6.22)$$

being the sum of two independent random variables, each of which is in turn the sum

$$\mu_n = \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_{ik}\xi_i\xi_k,$$
$$\nu_n = \sum_{i=1}^{n} \sum_{k=1}^{n} \beta_{ik}\zeta_i\zeta_k \qquad (6.23)$$

where ξ_i, ζ_i are normalized normal random variables. It has been shown that the random variables μ_n and ν_n may be represented as the sum of squares of n independent normal random variables. Then the probability distribution ρ_{2n} follows the generalized χ^2 distribution with $2n$ degrees of freedom. At $2n > 2$, one has to recourse to an approximate representation in terms of the Laguerre polynomials [162]. The average power of crosstalk for all the realizations along with the value which can be exceeded only in a given small percentage of realizations can be estimated. Several trial calculations may define the feeder requirements (their length and reflection coefficients).

The method of paired echoes used to determine crosstalk in an RF channel allows a probabilistic approach to be applied to the most important problem of determining the tolerances within which channel response may deviate from ideal one. This problem arises, for instance, in the inplant adjustment and checking of the equipment for a radio-relay link. In this case it should be determined which deviations of the channel characteristics (amplitude-frequency response and group delay time) from the ideal within a necessary bandwidth could be tolerable for one set of equipment (e.g., one transceiver) so that at the end of the link the probability for the crosstalk power to exceed a specified value be small.

The function $\mathcal{F}_B(\Omega)$ defined by Eq. (6.21) is assumed to be a random variable, as the samples of link characteristics at discrete frequencies are taken at random. The deviations of characteristics from the ideal within a required bandwidth are divided into systematic and random. Accordingly, the power of crosstalk also consists of deterministic and random components. Random deviations of each channel characteristics are caused by many independent factors, therefore they are distributed normally. It has been shown on those grounds that the crosstalk power is distributed by a generalized χ^2 law with many degrees of freedom. Under appropriate assumptions the relationships have been obtained, determining the mean crosstalk powers and their disposition at the receiving end of link in terms of tolerances for the amplitude-frequency response and the group delay time for one set of equipment. An approximate expression has been also derived for the probability of crosstalk power at the receiving end to exceed a given value.

6.3. Effect of Loading Type in FDM Multichannel Systems on Group Signal Amplitude Distribution

Problem statement. We consider now several issues associated with overloading in FDM multichannel systems. In these systems, independent messages are transmitted through a common channel by

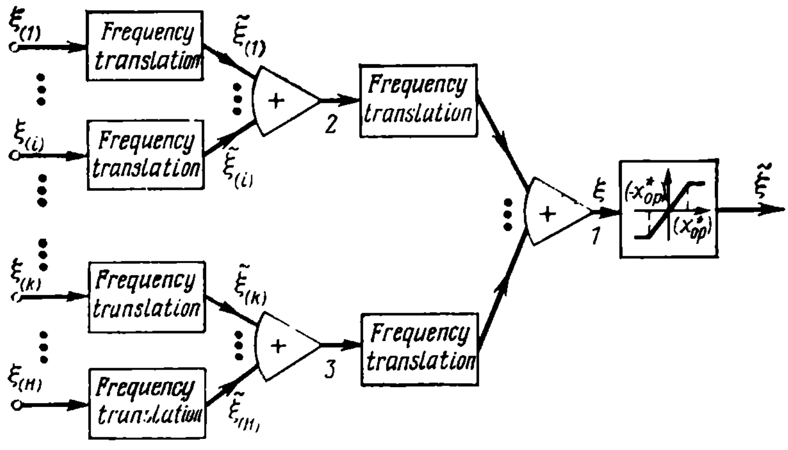

Fig. 6.2. Principle of frequency-division multiplexing system

way of translating each of them in different frequencies (Fig. 6.2). This translation might occur in several steps, with the signals being finally added up to give a group, or composite, signal. This composite signal (see 1, 2, and 3 in Fig. 6.2) is then transmitted through the different sections (e.g. amplifiers) of the communication link. At the receiving terminal, the group signal spectrum is separated into bands of individual messages by band-pass filters, and then, after retranslation of spectra initial messages are produced (Fig. 6.3).

At many transmission stages, this multichannel system is subject to distortions and interference. We now focus on one of the distortions caused by limited dynamic (amplitude) range of circuits handling the composite signal (see Fig. 6.3). In this operating range $(-x_{op}, x_{op})$, the nonlinearity of apparatus is, as a rule, insignificant. Beyond this range the nonlinearity is large. An overload is said to occur if the composite signal falls beyond the operating range, i.e. $|\xi| > x_{op}$. In real systems such an overload is inevitable. To a first approximation, the overload may be modeled by a pulse train opposite in sign to the truncated peaks of signals, as is shown by $n(t)$ in Fig. 6.3. This pulse train passes, in some form, through every band-pass filter of the receiver, and after the retranslation of spectra in frequency appears as an interference v. In particular, if speech signals are transmitted by the system, this interference manifests itself as clicks; for data transmission, this phenomenon is one of the causes of impulse noise.

In any event, overload impairs transmission quality. To a first approximation, the impairment may be characterized by the probability that the signal protrudes beyond the operating range (off-range probability). This probability is conditioned by several prac-

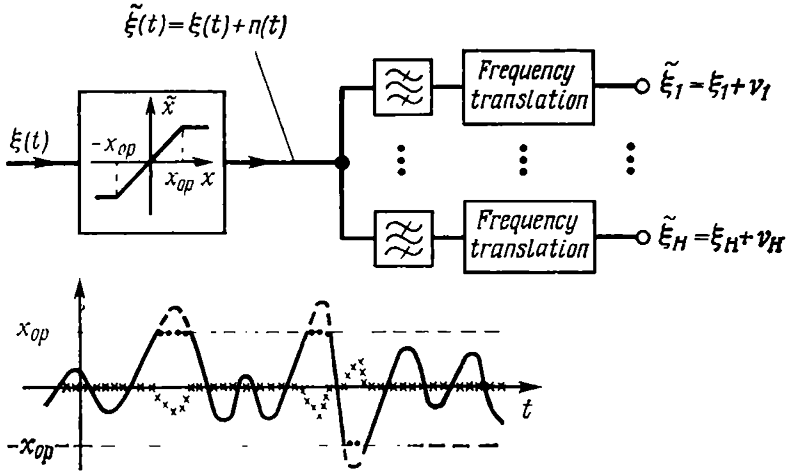

Fig. 6.3. Message recovery from a composite signal: — — — $\xi(t)$; $\cdots\cdot$ $\tilde{\xi}(t)$; ——— $\xi(t)$ or $\tilde{\xi}(t)$ if no overloading; × × ×—$n(t)$

tically important factors, primarily by the message type. It is common knowledge that most of FDM systems have been initially designed for speech signals. Then, however, they have been applied to low-frequency telegraphy, facsimile transmission, radio broadcast, and, in recent times, to data transmission. (These communications are known collectively as "secondary" loading.) The off-range probability is also dependent on the signal value, the combination in which individual messages join in the composite signal, and how many channels each of them occupies.

The design basis enabling the off-range probability to remain below a certain level has been developed earlier for the "uniform" speech load. (The uniformity here means that each channel is loaded with signals of one type, in our case speech signals.) These principles and rules are provided in the recommendations of CCITT. The situation has, however, changed of late materially. For one thing, in individual channels the load type changes with time; for the other, which is of more importance, the number of channels with "secondary" loads has increased.

Below, we shall discuss a method for the calculation of amplitude distribution for the nonuniform composite signal. The method is an improvement of "overload theory" due to Hollbrock and Dixon [389], and an extension to nonuniform group signal.

Overview of earlier results. An experimental investigation of uniform composite speech signals was first carried out in [357] and a more general treatment of these systems is associated with contributions of Hollbrock and Dixon [389].To give an outline of the results of these papers, we formulate the problem exactly.

The deterioration of transmission quality due to overload is related monotonously to the probability of the composite signal to exceed a level x_{op}.

$$P\{|\xi| > x_{op}\} = \Phi(x_{op}) \tag{6.24}$$

Therefore, the permissible value of quality deterioration may be put into correspondence with a so-called characteristic off-range probability ε. Notice that to one value of quality deterioration there correspond different values of the characteristic off-range probability. The overload problem can be regarded as solved if the off-range probability distribution function for the composite signal

$$\varepsilon = \Phi(x_\varepsilon) = P\{|\xi| > x_\varepsilon\} \tag{6.25}$$

or its inverse function

$$x_\varepsilon = \Phi^{-1}(\varepsilon) \tag{6.25a}$$

is known.

To be sure, the problem is also solved if we know the probability density function $w(x)$ for the composite signal, as

$$\varepsilon = \int_{-\infty}^{-x_\varepsilon} w(x)\,dx + \int_{x_\varepsilon}^{\infty} w(x)\,dx \tag{6.26}$$

In determination of these characteristics we assume that the composite signal may be written as

$$\xi(t) = \sum_{i=1}^{H} \tilde{\xi}_i(t) \tag{6.27}$$

$$\tilde{\xi}_i(t) = \sum_{j=-\infty}^{\infty} \tilde{\zeta}_{ik}\chi_{ij}(t) \tag{6.28}$$

Here $\tilde{\xi}_i(t)$ is the signal in the ith channel after translation in frequency; H is the number of channels in the composite signal; k is the message type (for speech one may think of two kinds of message: the speech proper and a break between phrases); $\tilde{\zeta}(t)$ are messages translated in frequency; j is the series number of a message; $\chi_{ij}(t)$ is the indicator of the presence of a given type message (Fig. 6.4). If messages in various channels are independent and the probability that kth message is in ith channel equals p_{ik}, then the following relationship may be written:

$$w(x) = S_{i=1}^{H}\{\sum_k p_{ik}\tilde{w}_k(z)\} \tag{6.29}$$

where S is the symbol of multiple convolution.

Hollbrock and Dixon assumed that in frequency translation (with no change in power) the message distribution does not change. (Theorems to this effect have been formulated in [378].) Denoting the

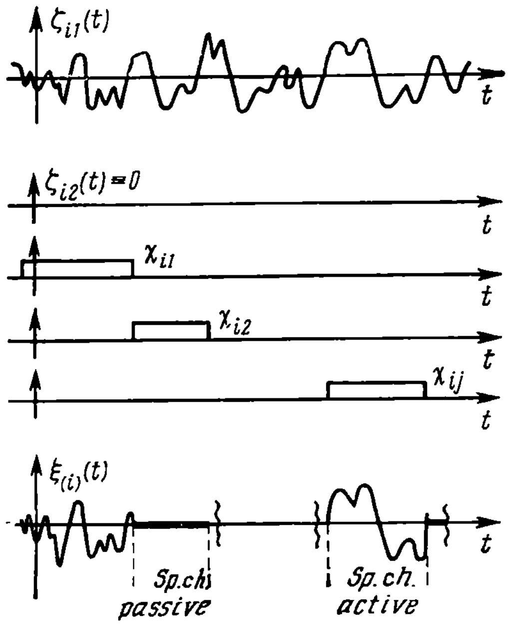

Fig. 6.4. Representation of speech message

message probability density by $w_k(z)$, we obtain the probability density of the composite signal as

$$w(x) = S_{i=1}^{H}\{\sum_k p_{ik} w_k(z)\} \tag{6.30}$$

In principle, Eq. (6.30) is the solution to the overload problem as it yields x_ε by using either Eq. (6.25) or Eq. (6.28).

These authors considered the case of homogeneous composite signal, which is equivalent to the analysis of the relationship

$$w(x) = S_{i=1}^{H}\{p_{sp} w_{sp}(z) + (1-p_{sp})\,\delta(z)\} \tag{6.31}$$

where the subscript "*sp*" means speech messages, and $\delta(z)$ is the delta-function indicating the probability density of breaks in speech. Equation (6.31) supposes that each channel is loaded with a speech message with equal probability, $p_{ik} = p_{sp}$. To determine $w(x)$ from Eq. (6.31), $w_{sp}(x)$ has to be obtained first [332, 348, 353, 357, 459]. The volume v was found to be determined by the relationship $\mathrm{E}\{\zeta^2\} = v$, and the distribution function $\zeta/\sqrt{v}$ was found to be independent of volume. This experimental evidence, according to [460], can be well modeled by the following Pearson distribution of the third kind:

$$w_{sp}(z/\sqrt{v}) = [1/\sqrt{v}\,2g\Gamma(l)]\,|z/g\sqrt{v}|^{l-1} \times \exp\{-|z/g\sqrt{v}|\} \tag{6.32}$$

here l is the microphone quality index taking on values between 0.5 (very good) and 0.2 (medium), $g = [l\,(l + 1)]^{-1/2}$. It has been established experimentally that the quantity

$$s = 10 \log v \tag{6.33}$$

obeys the normal distribution law with the mean $s_0 = -15$ dB and variance $D_s = 5$ to 8 dB. Hence, v has the lognormal distribution

$$w_v(y) = \left(h/D_s \sqrt{2\pi}\, y\right) \exp\{-0.5\,[(h \ln y - s_0)/D_s]^2\}$$
$$\text{for} \quad y \geqslant 0 \tag{6.34}$$

where $h = 10/(\ln 10)$.

The probability density for active speech is then

$$w_{sp}(z) = \int_0^{\infty} w_{sp}\left(z/\sqrt{y}\right) w_v(y)\, dy \tag{6.35}$$

It is obvious that the computation of Eqs. (6.32) and (6.33) and the multiple convolution in Eq. (6.31) is extremely cumbersome and it was not until the advent of computers that it could be performed. Hollbrock and Dixon have, therefore, proposed the following approach.

First, they assumed that the volume in each channel is identical and equal to v_0, where v_0 determines quantitatively the so-called short-term power of the composite signal

$$\gamma(t) = \frac{1}{T_1} \int_{t-T_1}^{t} \xi^2(t)\, dt \tag{6.36}$$

The operating range required for the composite signal was characterized in terms of a short-term power v_η and the probability of exceeding this level η. (The latter was taken to be about 0.01). This probability was supposed without proof to be connected with rms values of active speech signals, v_i, and with the probability p_m that at a given instant of time, m of H channels are active, by the relationship

$$\eta = \mathrm{P}\{\gamma > v_\eta\} = \sum_{m=1}^{H} p_m \mathrm{P}\left\{\sum_{i=1}^{m} v_i > v_\eta\right\} \tag{6.37}$$

The expression on the right hand side of Eq. (6.37) may be calculated knowing the probability density v_i. As a result, the parameter $v_{0.01}$ may be given as a function of the channel number approximately by the following form:

$$10 \log v_{0.01}(H) = -1 + 1 \log H, \quad 12 \leqslant H \leqslant 240 \tag{6.38}$$

The quantity v_0 is obtainable by the known value of v_η from the relationship

$$v_0 = v_\eta / pH \tag{6.39}$$

The probability density for the composite signal is thus

$$w(x) = \mathrm{S}_{i=1}^{H} \{ p w_{sp}(z/\sqrt{v_0}) + (1-p)\,\delta(z) \} \tag{6.40}$$

Yet even this expression has appeared too unwieldy for qualitative estimation. Hollbrock and Dixon have assumed, therefore, from physical arguments that for the composite signal the probability density has the form

$$\tilde{w}(x) = \mathrm{S}_{i=1}^{pH} \{ w_{sp}(z/\sqrt{v_0}) \} \tag{6.41}$$

For the required operating range expressed by

$$\varepsilon = \int_{-\infty}^{-\tilde{x}_{\varepsilon,\eta}} \tilde{w}(x)\,dx + \int_{\tilde{x}_{\varepsilon,\eta}}^{\infty} \tilde{w}(x)\,dx \tag{6.42}$$

it was assumed that

$$\tilde{x}_{\varepsilon,\eta} > x_{\varepsilon,\eta} \tag{6.43}$$

(The subscript η is introduced because the relationship $\mathrm{E}\{\tilde{\xi}^2\} = \mathrm{E}\{\xi^2\} = v_\eta$ holds.)

Now that the amount of information, other than speech, has increased and the extent of secondary loads has grown still further the overload issues once again have come to the forefront. Systematic studies in the field were under way since the 60s. The distribution functions for secondary messages have been derived. For music, an exponential distribution resulted

$$w_{mus}(z) = (1/\sigma\sqrt{2}) \exp(-\sqrt{2}\,|z/\sigma|) \tag{6.44}$$

and control signals and data transmission with AM and FM, PSK and on-off keying are well characterized by the probability density of the harmonic signal

$$w_{har}(z) = 1/\pi\sigma\sqrt{2-(z/\sigma)^2} \quad \text{at} \quad |z/\sigma| \leqslant \sqrt{2} \tag{6.45}$$

where $\sigma^2 = \mathrm{E}\{\xi^2\}$.

Computer-aided calculation of distribution function for group signal. Now we present some of the latest results obtained by one of the authors (Gordos, Hungary). A numerical method to compute Eq. (6.30) with inhomogeneous loading has been devised. Here we also assumed constant active speech power. To determine the multiple convolution, the fast Fourier transform was utilized, which appeared to be more advantageous than direct integration or the Monte-Carlo method [377]. This transformation is based on the relation

$$w(x) = \mathcal{F}\{\prod_{i=1}^{H} [\sum_k p_{ik} \mathcal{F}^{-1}\{w_k(z)\}]\} \tag{6.46}$$

Using further the definition of the characteristic function

$$\theta(y) = 2\pi\mathscr{F}^{-1}\{w(z)\} \tag{6.47}$$

we have, by Eq (6.46), the following expression:

$$w(x) = \mathscr{F}^{-1}\left\{\prod_{i=1}^{H}\left[\sum_{k} p_{ik}\theta_k(y)\right]\right\} \tag{6.48}$$

As a computational form, Eq. (6.48) is a most important intermediate result for analytical representation of characteristic functions of messages commonly met in practice. The value of these results resides in the fact that the program of fast Fourier transform is to operate only once, rather than $(k+1)$ times as Eq. (6.46) necessitates. The pertinent characteristic functions have the following forms:
for speech

$$\theta_{sp}(y) = \cos\left[l \arctan(g\sigma y)\right]\left[1+(g\sigma y)^2\right]^{-1/2} \tag{6.49}$$

for music

$$\theta_{mus}(y) = \left[1+0.5(\sigma y)^2\right]^{-1/2} \tag{6.50}$$

for a harmonic signal

$$\theta_{har}(y) = J_0\left(\sqrt{2}\,\sigma y\right) \tag{6.51}$$

where J_0 is the zero-order Bessel function of the first kind,
for a break

$$\theta_{br}(y) = 1 \tag{6.52}$$

for a normal process

$$\theta_{nor}(y) = \exp\left[-0.5(\sigma y)^2\right] \tag{6.53}$$

The inverse Fourier transformation $\mathscr{F}^{-1}$ in Eq. (6.48) is accomplished by using the fast Fourier transform technique. As a consequence, two errors arise. The first is associated with that the range of y should be bounded, as $\theta(y)$ can only be calculated on the finite interval $|y| < Y/2$. The upper bound of this error is

$$|\Delta_Y(x)| \leqslant 8\theta_0/|x| \tag{6.54}$$

where θ_0 is the maximum value of the ordinate lost because of the limitation. The second error is due to sampling, because the function $\theta(y)$ may be given only by a finite number $(N+1)$ of points. Besides the error, this involves another requirement, viz. for the interval $|x| \leqslant X/2$ the probability density may be defined only at those discrete points for which the relationship holds

$$XY = 4N\pi \tag{6.55}$$

Both errors might, however, be reduced materially by selecting N sufficiently large. Notice that if in each of the channels a break

occurs ($p_{i,\,br} \neq 0$), then θ_0 cannot be less than $\tilde{\theta}_0 = \prod_{i=1}^{H} p_{i,\,br}$. In order that this does not impair accuracy, the program uses the separation

$$\prod_{i=1}^{H} \left[\sum_{k} p_{ik}\theta_k(y)\right] = \theta^*(y) + \prod_{i=1}^{H} p_{i,\,br} \tag{6.56}$$

and the fast Fourier transform is applied to $\theta^*(y)$, with $\tilde{\theta}_0$ taken into account later.

The major inputs into the program to realize these principles are the probabilities that individual types of message occur in individual channels, p_{ik}, powers of individual messages, $v_{ik} = \sigma_{ik}^2$, and the number of channels, H. Additional inputs are N, X, and Y. The program yields, as needed, the characteristic function for the composite signal [see Eq. (6.56)], the probability density $w(x)$, and further, $\mathrm{E}\{\xi\}$, $\mathrm{E}\{\xi^2\}$, x_ε and some other parameters not mentioned here.

Limiting levels of secondary loads. Now we concern ourselves with the inhomogeneous composite signal containing other channels besides speech ones. With the above possibility of calculating the distribution function, we may proceed determining the levels of secondary loads. In so doing, it would pay to take into account that the quality of any speech channel in an inhomogeneous system should not be worse than that in the homogeneous one. This principle depends on the fact that the lower bounds of transmission quality (transmission and reception levels, signal-to-noise ratio, and so forth) cannot be appointed arbitrarily in the transmission of speech, as here the receiving end is a human ear. But if at least one end of the link is a device, then a lower quality is permitted with the less stringent tolerances allowed for the transmitter. receiver, and channel.

With speech, the quality of transmission is primarily decided by the intelligibility of syllables. From this viewpoint, group axis crossings (Fig. 6.5) are more harmful than individual ones. It is impossible to decide about the repetition rate of groups basing on the first kind distribution, however, this rate may be characterized in terms of the short-term average power γ exceeding over a certain level v_η. Therefore the required dynamic range is described by the parameter $x_{\varepsilon,\eta}$ to be determined in the same way as by Eq. (6.42), but subject to the condition that

$$\mathrm{E}\{\xi^2\} = v_\eta \tag{6.57}$$

In denoting the homogeneous composite signal by ξ and the inhomogeneous by ξ^*, one should keep in mind the relation

$$x_{\varepsilon,\,\eta}^* \leqslant x_{\varepsilon,\,\eta} \tag{6.58}$$

It is easily seen from what has been said that $x_{\varepsilon,\,\eta}$, $x_{\varepsilon,\,\eta}^*$ are dependent on the total number of channels, on the number of chan-

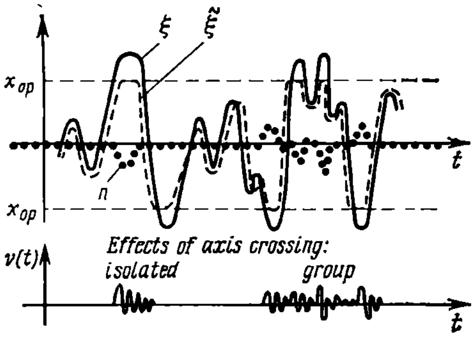

Fig. 6.5. Individual and group axis crossings

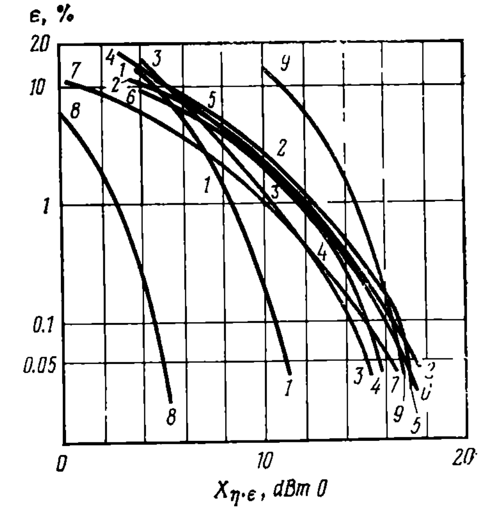

Fig. 6.6. Off-range probability for 12-channel group signals

nels with secondary load, their power and probability of presence. This is illustrated by the following relation for a simple case where the group signal consists of speech and a secondary load of one type

$$\Phi_*^{-1}\{\varepsilon, \eta, v, p, v^*, p^*, h^*, H\} \leqslant \Phi^{-1}\{\varepsilon, \eta, v, p, H\} \quad (6.59)$$

(The asterisk here indicates the secondary load, and h^* is equal to the number of channels with the secondary load.) Eq. (6.57) in that case is written as

$$h^*v^* + (H - h^*)\, v = v_\eta \quad (6.60)$$

where v_η is defined by Eq. (6.37). If the probability of secondary messages presence is unity ($p^* = 1$), then this relationship changes to the following:

$$\eta = \mathrm{P}\{\gamma > v_\eta\} = \sum_{m=1}^{H-h^*} p_m \mathrm{P}\left\{\sum_{i=1}^{m} v_i + h^*v^* > v_\eta\right\} \quad (6.61)$$

which only means that in Eq. (6.38) $(H - h^*)$ should be substituted for H, and $(v_\eta - h^*v^*)$ for v_η. Solving Eq. (6.59) for v^* may give the permissible level of secondary loads.

Some practical conclusions. The results of off-range probability studies for several composite signals in a 12-channel communication system are given in Fig. 6.6. In channels with various types of communication, the levels were selected, as a rule, according to the recommendations of CCITT, but in data transmission channels higher levels were set. The examination of the curves in Fig. 6.6 and the data in Table 6.1 warrant the following conclusions.

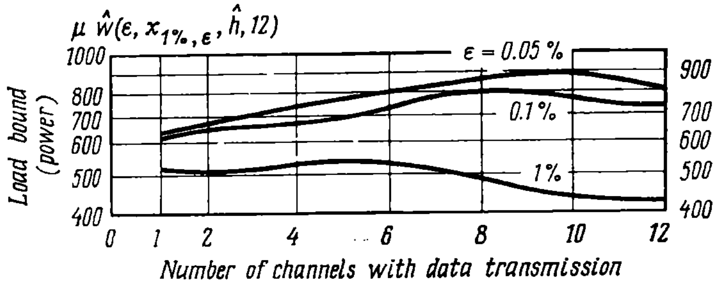

Fig. 6.7. Limiting power load for data transmission channels of 12-channel group

The required operating range for the homogeneous signal may be reduced with better microphones (see curves *2* and *3*).

The levels of secondary messages, safe for data, are selected so that the required operating range for inhomogeneous composite signals is essentially the same as for homogeneous.

TABLE 6.1

Signal	Composition of group signal	Power of group signal	
		mW0	dBm0
1	Normal process with power as in *2* and *3*	2.136	3.296
2	12 speech channels $l=0.2$	2.136	3.296
3	12 speech channels $l=0.5$	2.136	3.296
4	9 speech channels, 1 for music	2.352	3.714
5	11 speech channels, 1 facsimile	2.958	4.709
6	11 speech channels, 1 telegraph	2.093	3.207
7	6 speech channels, 6 for data	1.369	1.360
8	12 channels for data transmission	0.6	—2.218
9	12 channels for data transmission	9	9.542

In overload measurements, it would suffice to simulate only speech channels (in measuring at the level of channels) and the homogeneous group signal (at the level of group). Thus, in measuring the principle of "common noise loading" may be adopted.

The dynamic range of the composite signal appears to be smaller than that recommended by CCITT for the homogeneous load in a 12-channel system.

The last conclusion prompted calculations to determine the upper bound of data signal levels, for which the required operating range

of an inhomogeneous composite signal is the same as for a homogeneous composite signal (Fig. 6.7). The curves were plotted assuming that the peak values of probability ε pertaining to a short-term average power exceeded with probability η coincide in the homogeneous and inhomogeneous groups:

$$x_{\varepsilon,\eta} = x^*_{\varepsilon,\eta}$$

It is readily seen that the permissible level depends on the quality required. Obviously, the value 50 μW0 contained in CCITT recommendations may be increased at least by 10 dB for a 12-channel system without any quality impairment. This fact is of paramount importance for systems of data transmission over leased lines.

Chapter

7

Communication Systems

7.1. Modern Methods of Transmission and Reception in Satellite Communication Systems

Satellite communication systems. Transponders of communication satellites usually operate within the frequency bands allocated by the Radio Regulations. They are about 6 GHz for the earth to satellite direction of transmission and about 4 GHz for the satellite to earth direction. New ranges of 11 and 14 GHz are bringing into service. The total passband of a satellite station is usually 500 to 700 MHz. It is divided into 35-40 MHz bands with amplification performed by separate transponders. The output power of a transponder is usually 5 to 15 W, reaching 50 to 200 W in TV broadcasting systems. The modern space technology ensures high accuracy of orbiting and orientation of a satellite (up to 0.1°), such that spot-beam antennas can be employed with beam width from 17° (global coverage) to 2-4°. Earth stations (ES) of satellite communication systems (SCS) are provided with parabolic antennas of various diameters (2 to 32 m) and low-noise receivers. The ES figure of merit (the ratio of the antenna gain, G, divided by the system noise temperature, T) attains 41 dB/K.

Major sources of noise in satellite communications are the input stages of the ES receiver, account should be also made of noises due to the input circuits of satellite receiver, antenna waveguides, the Earth and other sources of cosmic noise. All these sources produce the noise of thermal type. With the above energy parameters the signal to thermal noise ratio at the ES demodulator input, as measured in the RF passband of transponder, ranges from 8 to 20 dB and more. This allows each transponder to accommodate from 200 to 800 telephone circuits or one or two television programmes, and in some systems up to 1900 voice channels.

Satellite communications over large distances (about 1-5 thousand kilometers) appear more attractive economically than terrestrial links. Yet especially attractive the satellite transmission proved for

conference (multi-destination) messages (TV and radio broadcasts, newspaper facsimile) beamed down a wide network of earth stations, and for telephony over difficult or wide water terrains or sparsely populated regions [256, 269].

Satellite telecasts. For TV messages, frequency modulation in the analog form (FM) is used. The immunity of this communication technique to the action of thermal noise with conventional demodulation procedure employed is well studied (see, e.g. [124]). The evaluation of threshold properties, however, should make allowance for the structure of colour television message and subjective perception [157]. The analysis of threshold in TV reception uses the model worked out by Rice (its detailed description is given in [124]). The model is based on the conception that the threshold is caused by $\pm 2\pi$ phase jumps in received signals, which occur when the signal, u_s, and noise, u_n, are added together. At the output of the frequency detector (with unit slope) at the time of a phase jump, a voltage pulse arises with an area of 2π.

To decrease the low-frequency portion of the video signal spectrum, use is normally made of linear pre-emphasis standardized by CCIR. It reduces materially the non-linear distortions of the message, which are evaluated by the standard two-frequency test signal. An analysis of FM signal distortions is presented in [35]. Linear pre-emphasis has no influence on the communication link S/N ratio (measuring noise above threshold with a weighting network).

Non-linear pre-emphasis has also found some use [270]. Essentially it is a limitation of overshoots of a signal after linear pre-emphasis followed by an increase in its peak-to-peak value. These overshoots arise if in the initial message brightness of large amplitude sharply changes its value. The probability of these changes is small, thus making them amenable to limitation without any strict requirements on the accuracy in de-emphasis.

Great efforts went into the development of tracking demodulators having improved threshold properties. But the results attained are rather modest. The reason is to be found readily from the impulsed clicked model of threshold of tracking demodulators [124]. The model basically relies on the fact that the tracking demodulator reproduces those of $\pm 2\pi$ phase jumps coming to its input which it manages to track, and suppresses (does not feel) the jumps which it fails to track. To gain sufficiently in threshold, therefore, it is necessary to decrease the amount and bandpass of feedback in the tracking loop. Yet the useful modulation of the signal should be tracked with sufficient accuracy to avoid message distortions and tracking disruptions which give rise to effects similar to threshold pulses. The television message contains large, short and sharp-edge overshoots aggravated by linear pre-emphasis. To be tracked fairly accurately the overshoots require such a wideband feedback that

the input phase jumps caused by noise are well tracked and the tracking demodulator approaches the conventional one in threshold. It is to be noted that a television signal is associated with radio broadcasting or newspaper facsimiles, their transmission being on subcarrier frequencies above the television signal spectrum (6 MHz), which also calls for wideband tracking.

The immediate perspectives of satellite television broadcasting systems are associated with technological improvement of the communication channel rather than with the message processing. Television systems are being developed to work with relatively simple receiving stations, for which purpose the antenna gain and power of the satellite station are increased (within the limits allowed by the Radio Regulations).

Television-associated sound program. With frequency-division multiplex, a modulated subcarrier frequency of 6.5-7.5 MHz is used. In professional (high-quality) reception at large earth stations, a subcarrier frequency of 7 MHz and higher is utilized with an optimum FM index ($m \approx$ 4-5); use is also made of the tracking subcarrier demodulator and even controlled companding, which makes it possible to spend on the TV sound signal less than 0.5 to 1 dB of the energy resources of television channel. Satellite mass distribution systems (e.g., *Ekran* system) use a subcarrier frequency of 6.5 MHz with a frequency deviation of 50 kHz, which simplifies the forming of a standard TV signal for terrestrial-broadcasting at the receiving ES (for this signal, the difference between video and sound carriers frequencies is 6.5 MHz with frequency deviation in the sound channel 50 kHz). This form of high-fidelity sound transmission causes energy losses of 2-3 dB in the video channel. Sound transmission with subcarrier frequency also requires better linearity of all the radio channels for the message being transmitted, primarily to avoid amplitude cross modulation, and hence the appearance of a threshold in the subcarrier channel.

The Soviet-made distribution television system *Orbita* uses the TDM technique relaying mainly on the so-called structural redundancy of the standard television signal. In that case, at the intervals of the line blanking pulses, one or two sound channels are transmitted in a discrete form. In the *Orbita* system two-side pulse-duration modulation [241] is used for the purpose and the transmission of either one sound channel with bandwidth 10 kHz or two separate channels with bandwidth 6 kHz each is provided. A high signal-to-noise ratio in the video channel (about 30 dB) enables multilevel PCM to be utilized for the same purpose. With this modulation up to five or six 15 kHz channels may be transmitted within the intervals of line blanking pulses. The TDM technique applied to the television channel of main communication links offers much promise and its potentialities are far from being exhausted. A further increase in

the number of sound channels may be achieved with the television channel time-division multiplexed in the IF path. This permits of a capacity of 2 to 5 Mbit/s, being equivalent to 8 to 20 15-kHz channels.

Telephony. Satellite transmission of telephone messages has an important feature differing them from terrestrial links—so-called multistation (or multiple) access. This term denotes the way in which a communication satellite may simultaneously work with several ESs within coverage area serving them through the common on-board transponder circuit.

Most often the frequency division multiple access (FDMA) is employed with each carrier frequency modulated by frequency division multiplexed message. When selecting optimal parameters (frequency deviation at each carrier, power utilization of satellite transponders), statistical properties of multichannel telephone traffic and threshold characteristics of the receiver are to be taken into account. To lower the threshold of demodulator a synchronous phase detector is normally employed, that operates without threshold "clicks" at an input S/N ratio of about 8 dB. As follows from the analysis [101] with a correct selection of parameters, the losses due to multiple access provision are not significant. Capacity losses are characterized by the ratio of the peak power of a transponder required to transmit N signals to the power required to transmit the same total number of channels with one carrier. The choice of the optimal frequency deviation, however, is possible with a relatively small energy potential of the RF link. With substantial energy reserves (e.g. in using spot-beam satellite antennas), the limiting factor becomes the RF channel passband, therefore, the frequency deviation has to be reduced, thus bringing about an increase in multiple access losses. As is indicated in [101], the losses increase also with the number of FM signals transmitted via the common transponder. In transmitting each telephone channel on an individual carrier (SCPC—single channel per carrier), however, a possibility arises to more than double the transponder capacity by suppressing the carrier during the telephone conversation pauses. In the process, at each carrier either analog FM or PSK and PCM is realized, each channel rate being 64 kbit/s. The latter technique has some advantage only when working with ESs having large-size antennas.

One more way to achieve a higher capacity with SCPC is the channel demand assignment (DA). Here the channels are not preassigned to one or another ES but are provided on request. With a sufficiently low probability of refusals to provide a channel (about 1 per cent), the transponder may handle 2-4 times more messages [124].

With further development and improvement of methods for signal transmission in a discrete form, and construction of efficient PSK modulators and demodulators, a trend appears toward satellite

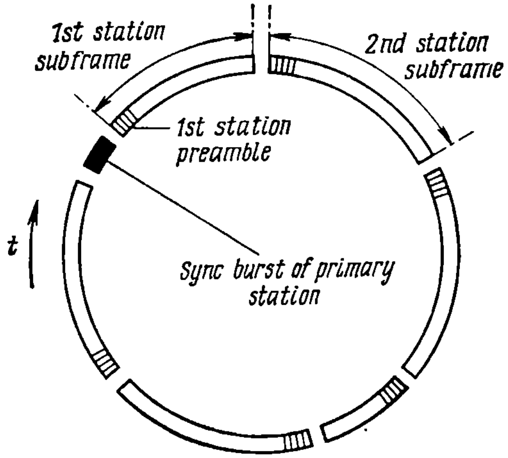

Fig. 7.1. TDMA frame

communication systems in which signals are time-division multiplexed by on-board transponder (TDMA) [394, 395]. In such a network each earth station is allocated to radiate its signals within periodically recurring time slots whose duration in the general case is defined by the station traffic. The time interval during which all the stations in the network radiate their signals once is called the frame, and the length of the burst radiated by one station the subframe (Fig. 7.1). Such a system enables the transponder to be used as if in a single-carrier mode, more exactly, in that of maximum power. System capacity losses due to guard time between the subframes are less than 1 per cent, and those due to preamble bursts, less than 10 per cent, even for a large number (several dozens) of stations in a network.

In this case, however, in transmitting wideband PSK signals there arise characteristic intersymbol distortions or interchannel distortions if four-phase PSK is used. Attempt at their minimization imposes much more stringent requirements upon amplitude-frequency and phase characteristics of radio circuits than with FDMA [336].

Direct analytical investigation into what effect RF circuit characteristics of a communication link may have on the FM signal distortions and associated energy losses poses severe mathematical obstacles due to description of random signals and a requirement to represent the RF circuit characteristics in a sufficiently simple form. In recent years a trend has become apparent to apply computers to solve the problem. Some latest works in this field estimate the distortions in real RF channels to various degrees of accuracy.

In developing satellite TDMA communication systems a challenging engineering problem is that of mutual synchronization of stations to initially pull a station into synchronism and maintain a subframe at a specified time slot. In TDMA satellite communications, as a rule, the operation of time companding or transforming of binary flow rate is required. Signals in discrete form are thus particularly convenient for such systems, because with binary signals the storage is readily obtainable with integrated circuits, and it is this storage that forms the basis for time companders.

Telephone signals are mostly analog. The conversion of these signals into digital form is a separate problem that has several different solutions. The widest acceptance has been gained by the PCM technique improved by the use of digital instantaneous companders. Here PCM converters are used to process both individual telephone channels and standard group (e.g. 12- and 60-channel) signals.

Also of interest are the recently devised methods of analog-to-digital (A-D) conversion using adaptive PCM and delta-modulation (DM); these techniques permit of a reduction in the output digital flow rate and a large probability of error in reception (see Sec. 7.2). The current A-D converters ensure high quality of telephone messages transmitted at a rate of about 64 kBaud per channel with PCM, and 32 kBaud per channel with DM. But when telephone channels are used for telegraphy, e.g. facsimile, data, etc., the rate of DM digital flow should be brought up to 56-64 kBaud per channel.

For digital signals, the satellite communication uses normally PSK, and depending on the energy aspect of a radio link and allocated bandwidth, two-phase or four-phase PSK may be utilized. For the current satellite communication parameters, the total RF transponder capacity ranges from 20 to 60 Mbit/s. The transmission of digital streams at rates about 40-60 Mbit/s is the most attractive form economically, as it allows a tradeoff in the current satellite RF circuits between the power aspects of the radio link and the bandwidth.

Radio broadcasting. In transmitting radio broadcast programs by common telephone FDM systems, each radio broadcasting channel appears to be equivalent to ten or more telephone channels. This comes about because radio broadcasting communications have a wider dynamic range, and in addition, in national systems, coinciding programs may be transmitted via many channels.

If each program is transmitted by individual carrier (with FM or FSK-PCM), the transmission of broadcasting programs turns out to be not so efficient because of crosstalk caused by the non-linearity of satellite transponder circuits, leading to 3-5 dB power loss.

With a large number of broadcasting programs, the most efficient is TDM technique. Stringent requirements to be met by broadcast

transmission, especially of the first and extra classes, necessitate a marked increase in the pertinent A-D conversion rate. This is accounted for by a low level of quantization noise required, a distortion type characteristic of A-D conversion. To this end. 11 to 14-digit PCM converters have been developed. The companding of dynamic range of signals enables these multidigit code words to be transformed without loss of quality into 8-10 bit combinations prior to their feeding to radio channel. This transformation is only possible because the quantization noise is most pronounced with low-level signals. Therefore it is reasonable for the quantization step to be increased at high levels and retained at low levels. The operation might be accomplished either at the A-D conversion unit input by means of a non-linear amplitude characteristic or in a digital form at the A-D converter output by using recoding devices. Thus, a first-class broadcasting channel may be organized using a digital flow with a rate of 200-250 kbit/s, and an extra-class channel with a rate about 350 kbit/s.

It should be noted that the discrete form of signals in telephony and broadcasting enables them to be transmitted simultaneously in one transponder. Accordingly each broadcast channel can be replaced by a number of telephone circuits, which in contrast to FDM is only dictated by the rates of pertinent bit flows.

An estimation of potentialities for the capacity growth in satellite telephony warrants a conclusion that the greatest effect is to be expected from higher-energy-density communication channels. They should be based on narrow beams properly stabilized in space and switched by an appropriate algorithm determined by the multiaccess technique employed. Another trend is to increase the number of satellite transponders (those with polarized decoupling included) and the use of signal processing by the satellite.

With telephone signals in a discrete form, a substantial increase in capacity can be attained by a statistical multiplexing. This technique uses natural pauses in telephone conversations (see Sec. 1.7) enabling the number of transmitted channels to be increased without an increase in the rate of the total bit stream. With statistical multiplexing, the telephone network capacity can be more than doubled by digital speech interpolation (see Fig. 7.2 taken from [433]).

Furthermore, a promising method for improving the capacity of transponder and fidelity of reception is forward error correction (see Chap. 3). The S/N improvement ensured with their help may be as high as 5 to 7 dB.

In estimating the advantages and disadvantages of various transmission techniques one of the most important properties is their immunity to radio interference in coinciding and adjacent channels. With this objective in mind, one should consider both the electromagnetic compatibility with earth radio stations (see Sec. 7.4) and

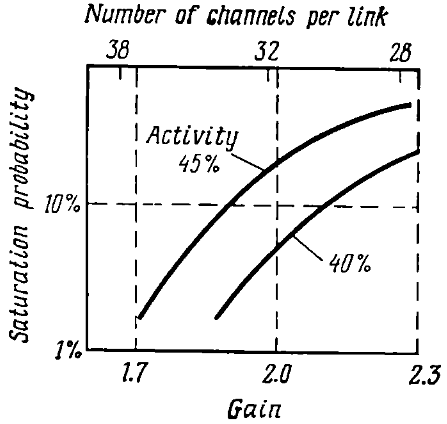

Fig. 7.2. Increase in capacity with digital speech interpolation (the number of conversations transmitted is 64)

efficient usage of the geosynchronous orbit by the set of satellite communication systems.

Geosynchronous orbits. By mid-1977, CCIR has registered about 100 communication satellites in a geosynchronous, or geostationary, orbit. Being unique in its properties, the orbit has found such a wide use that any insertion of new satellites, especially operating in the common 4 and 6 GHz ranges, not infrequently is difficult or even impossible because of interference with existing systems. This concerns first of all some of the most convenient sections of the orbit, e.g. over the Atlantic and Indian oceans.

With signal and equipment parameters common for satellite communications, the angle separation between neighbouring satellites should not be less than 4 to 6°, though at times it may be reduced at least by half.

To keep the interference between the adjacent systems to a minimum, various means may be used, viz:

improved spatial selectivity of satellite and earth station antennas;

minimized areas covered so that to be confined by the region being serviced, or even by small "spots" within the region;

agreed frequency plans of adjacent satellite systems;

high stability of satellite orbiting position and high accuracy of satellite antenna pointing;

close energy parameters of adjacent satellites; error correction methods of signal transmission and reception;

signal energy dispersion over the spectrum using appropriate signals;

standardized interference between adjacent systems to be increased from 1000 to 2000-3000 pW in a channel.

Given a permissible noise level in a channel (10 000 pW), heavier interference from adjacent systems calls for a reduction in capacity of each separate system. It has been found nonetheless [123] that these losses are more than compensated by the possibility to place into the orbit a larger number of satellites, with the mentioned values of crosstalk corresponding to the maximum orbit capacity.

In this connection, the quantification of efficiency with which the geosynchronous orbit is exploited by one or another satellite communication system comes to the forefront. It is most adequate to estimate the satellite communication system efficiency in terms of channel-kilometers of equivalent terrestrial network [123]. The orbit arc section occupied by this system is defined by a coefficient D taking into account that part of the earth's territory, S_E, which cannot be used should the system be located at the point on the arc with angle coordinate ϑ:

$$\bar{D}=\frac{1}{2\vartheta_{\max}}\int_{-\vartheta_{\max}}^{\vartheta_{\max}} D(\vartheta)\,d\vartheta$$

where $D = S_E/S_0$; S_0 is the area on the earth's surface covered from the geosynchronous orbit; $\pm\ \vartheta_{\max}$ are the arc boundaries of the orbit occupied by the satellite communication system $[D\ (\theta > \theta_{\max}) = 0]$.

The occupied orbit arc should be evaluated as related to a reference standard system whose parameters should be worked out and agreed internationally.

7.2. Adaptive Discretization

Algorithms. In digital processing of analog signals, on-line signals input into computer or in transmitting through digital communication channels, an analog-to-digital (A-D) conversion is needed, that involves time sampling and amplitude quantization of an incoming signal. In sampling at times t_k, the samples $x_k = x\ (t_k)$ of the signal $x\ (t)$ are formed. The quantizer is characterized by the quantization range 2Δ spanned symmetrically about the reference level. The signal dynamic range is broken down by the quantization levels which are usually equidistant with quantization step h, into quantization intervals numbered, for instance, in increasing order of magnitude. If the analog signal x_k falls within the jth interval, the quantized quantity is assigned the same value $\hat{x}_j$. Should the analog sample fall beyond the range (over or below), the quantized quantity is given the maximum or minimum value, respectively.

In A-D conversion, the major efficiency indices are the value of error in recovering the initial analog signal from digital data and the average message formation rate measured in terms of bits issued per unit time. For the stationary random input signal with known statistical parameters, the separation between the sample, providing the minimal rate for a predetermined error, is decided by the bandwidth. In that case the quantization level coincides with the average, the bandwidth is found from the variance value, and the quantization step from the known distribution [90, 280, 312, 436]. Usually the reference level, and also the average value of the signal, is assumed to be zero.

The model of the stationary random process is an approximation and is only suitable for a qualitative description of real signals. Under nonstationary conditions or a priori uncertainty, the adaptive A-D conversion techniques are proved to be efficient. The adaptation presupposes the current evaluation of the bandwidth, variance, mean, and distribution function of the signal, and also the adjustment of the A-D conversion parameters in accordance with the information obtained. In order that the analog signal may be recovered, the discrete data should be supported with an additional information, about those values of changeable parameters of A-D conversion for which these data have been obtained. Two known methods for constructing the adaptive A-D conversion differ in the way by which this additional information is derived.

In the first method, the analog signal is applied to the input, i.e. the measuring device is connected to the input of the A-D unit. The values of controlled variables should be transmitted through the communication channel as additional information [12, 203, 240].

The second method does not presuppose any transmission of auxiliary information. here the data on the signal parameters are derived at the output, i.e. from the discrete data available both at the receiving and sending end of the communication channel. We shall use a common name for these algorithms: adaptive pulse-code modulation (APCM). It is desirable to distinguish the APCM systems by the ensembles of changeable A-D parameters [211].

Examples. Classed with one-parameter systems are delta-modulation (DM) and differential pulse-code modulation (DPCM) systems. Use is made of the elementary binary quantization of the difference ε_k of the current analog sample x_k and preceding quantized value $\hat{x}_{k-1}$. The estimate is determined by the step h

$$\varepsilon_k = x_k - 0_k, \quad \hat{x}_k = x_{k-1} + \hat{\varepsilon}_k \tag{7.1}$$

$$\hat{\varepsilon}_k = h \operatorname{sgn} \varepsilon_k; \quad 0_k = \hat{x}_{k-1} \tag{7.2}$$

$$\operatorname{sgn} \varepsilon_k = \begin{cases} 1, & z > 0 \\ -1, & z < 0 \end{cases} \tag{7.3}$$

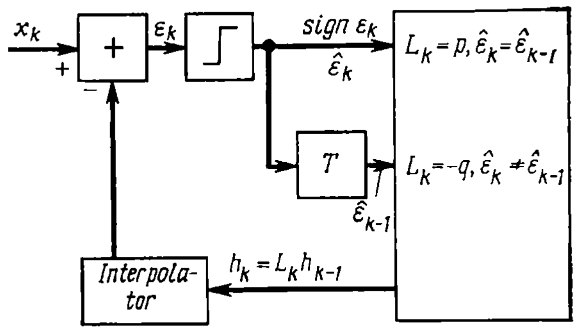

Fig. 7.3. Schemes of the Jayant algorithm

The DPCM scheme differs from the delta-modulator in that it uses 2^L quantization intervals and stepwise response of the quantizer.

In order to expand the possible amplitude range of the input signal, the DM step is adjusted. These forms of two-parameter systems include DM with double integration, which provides the linear growth of step to catch up with a sharp change of the input signal, and its modifications (various instantaneous companding circuits, and high-informative adaptive delta modulation [469]). The subject is discussed at length in [71, 187, 346, 392, 487].

As an illustration, consider the Jayant algorithm [392] depicted in Fig. 7.3. Each next quantization step, h_k, is obtained from the preceding one according to the relationship

$$h_k = L_k h_{k-1} \tag{7.4}$$

and the factor L_k is obtained from the current and preceding output processes in the quantizer: if their signs are similar, then $L_k = p$, and with unlike signs $L_k = q$, where $p > 1$ and $q < 1$. It is suggested in [383] to change the spacing between the samples along the similar lines. When this algorithm is supplemented by the step adjustment, we have a three-parameter system [383]. The counterpart of DM with an alternative step is a three-parameter system—DPCM with controlled range and such a step variation for which the number of levels in the range remains constant [392]. The use of the "matched" step variation and range to retain the number of digits in a conventional PCM scheme results in an adaptive quantizer [443]. The algorithms are also suggested for the variation of arrangement of quantization levels pertaining to the three-parameter system, the so-called switched quantizer.

Modeling and experiment. The adaptive A-D systems are suitable for communication technology to transmit speech and picture signals. They have gained in importance in recent years in connection

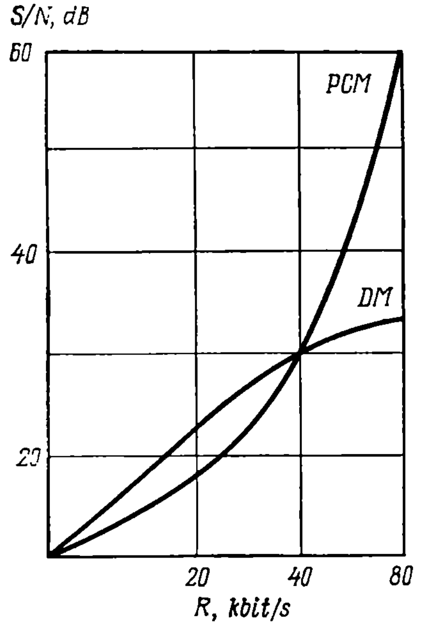

Fig. 7.4. Signal-to-noise ratio vs. bit rate

with the advent of digital transmission techniques. In the early stage of their introduction, a modulation method was needed that compared favourably with the analog systems in its cost effectiveness [229]. Thus, PCM was preferred (Fig. 7.4).

Several ways are known to test APCM. Earlier studies relied on mock-ups, and therefore used the harmonic analysis. The results obtained in references [330, 351, 469, 490] are summarized in Fig. 7.5 as a variation of the signal-to-noise ratio with the input amplitude at various frequencies. The experimental conditions were about similar: noise was measured within a bandwidth of the order of 2.5 kHz, the message rate being about 50 kbit/s. From the data of Fig. 7.5, Table 7.1 is derived, listing signal-to-noise ratio and dynamic range (width of maxima at 5 dB) calculated from four frequen-

TABLE 7.1

DM type	S/N, dB	Dynamic bandwidth, dB
Conventional [490]	38	10
IBM make [469]	37	26
With simple logic [330]	26	17
With 2/3 companding [490]	20	28

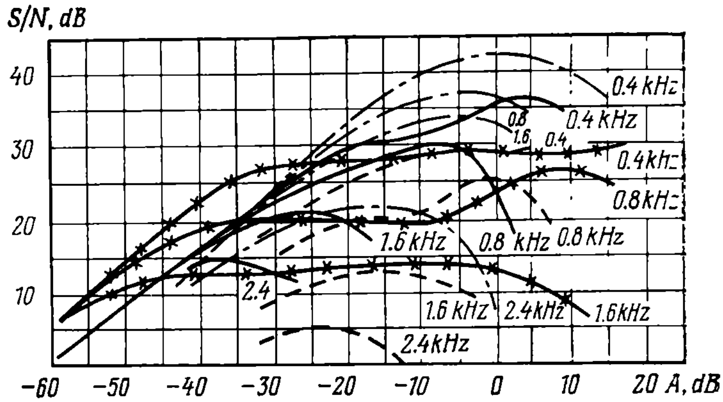

Fig. 7.5. Variation of S/N ratio with input amplitude: ———— delta modulator [330] with simple logic of step changing; — — — delta-modulator [490]; —·—·—· IBM delta-modulator with companding; —x—x— DPCM with variable step

cies of the input signal: 400, 800, 1600, and 2400 kHz. The table illustrates the marked gain provided by the changing over to alternative-step systems.

In using a random input signal, the algorithm is modeled with a computer. Figure 7.6 shows the variations of the signal-to-noise ratio with the variance of the Markovian signal for a correlation coefficient between adjacent samples of $\rho = 0.5$. The upper curves refer to the Jayant algorithm (see Fig. 7.3) at different values of p and q [509]. The lower curves characterize DPCM with variable step [393].

An important parameter of the system is the ratio R of maximum to minimum permissible steps. The dotted lines relate to pertinent systems with an invariable step. It follows from Fig. 7.6 that the dynamic range essentially coincides with the value of R. The optimal values of p and q for the Jayant algorithm lie in intervals: $p = 1.1\text{-}1.3$, $q = 0.85\text{-}0.95$.

Figure 7.7 shows variations of S/N ratio with the variance of the input signal for two forms of APCM and the random signal [407]. The lower curves relate to delta-modulation with an alternative and a constant step at a sampling frequency of 40 kHz, the upper to DH system (PCM with variable step) with the same pattern of step variation, and to PCM systems with constant parameters at a sampling frequency of 56 kHz.

The most complete insight into the APCM algorithms may be gained by subjective tests. In speech communication, the method of equiprobability curves is applied [356], that is those combinations of parameters are found for which the performance quality is identical.

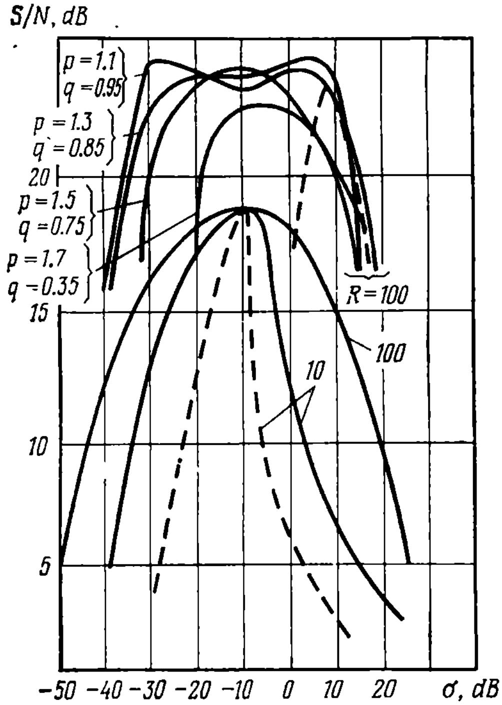

Fig. 7.6. Variation of S/N ratio with variance of the Markovian signal

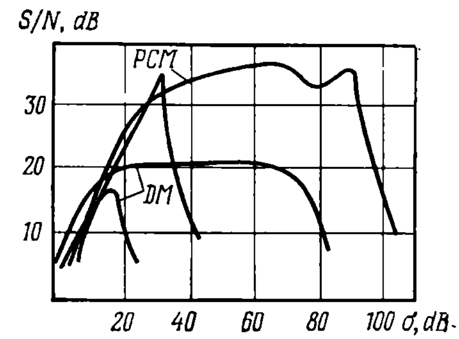

Fig. 7.7. Variation of S/N ratio with variance of input signal

It is shown in [393] that in speech transmission the changing over from the optimal PCM to DH system results in a 4 to 8 dB improvement in the S/N ratio. The incorporation of the reference level adjustment resulting in the DOH system (DPCM with a controlled range and matched step variation) improves the S/N ratio by another 3-4 dB so that the resulting improvement due to the adaptation is as high as 10 dB. The delta modulation with variable step ensures double reduction in capacity as compared with PCM in a telephone channel.

In choosing the APCM parameter under real conditions, there are no all-purpose test signal and quality criterion. Therefore, a multi-step test procedure might be recommended. In the first step, simple methods of harmonic analysis are applied with root-mean-square criterion (e.g. S/N ratio) to effect the primary selection of algorithms. Further, a random signal may be used with the same criterion. Real signal studies are more laborious. The final stage is the full-scale tests which normally are rather involved and costly. Therefore, the number of algorithms selected for such tests from the tentative analysis should be small.

7.3. Fading Control in Analog Data Transmission

Introductory remarks. Fading in a multipath communication channel is a cause of multiplicative interference. In the presence of additive noise it is the major obstacle to reliable signal transmission. Therefore, the development of methods for fading control is a fundamental problem underlying the feasibility to transmit signals over the multipath channels.

Multipath channels have been first employed to transmit information in the 1930s when space-diversity reception was used to fight fading at short waves. During the years which followed, other diversity methods have been proposed along with techniques to combine diversity signals in the high frequency channel [186, 455]. Fundamentally new fading control techniques are associated with the discovery in the early 50s of the troposcatter long-distance VHF propagation and the utilization of that effect to establish wideband radio links [333]. Fading appears in communication channels with VHF ionoscatter, using the reflection of radio waves from a man-made layer or the Moon, and in optical channels. Though physical processes underlying the multipath nature of all the above channels are different, their description may be reduced to a certain generalized model in which the parameters alone are to be changed.

Fluctuations in the level of a received signal are a nonstationary random process which is a function of absorption, scattering and reflection of electromagnetic waves in the propagation path. Taken over limited time intervals, however, this process is, as a rule, "locally" stationary, to facilitate the construction of sufficiently simple mathematical models of the channel (see Sec. 1.4).

In the general case the transfer function of the channel is random in both time and frequency, $K(t, f)$. In most cases when evaluating how advantageous in signal amplitude distribution a fading control system is, the signal is assumed to be narrow-band[1], and fading smooth. Should such systems also reduce waveform distortions, then the evaluation is made on the assumption that the signal level is constant (with the exception of calculations of AM-PM conversions in the receiver limiter). Physically, the conditions for fading to be deemed smooth are defined by the product of channel response time (memory), $\Delta\tau_{max}$ to its bandwidth, $2\Delta f$. If $\Delta\tau_{max}\,\Delta f \ll 1$, the fading may be regarded as being smooth.

Analytically, the nonuniformity of the frequency response may be evaluated in terms of the correlation coefficient for the fluctua-

---

[1] The signal narrowband nature in a multipath channel in which all the spectral components of the signal fade simultaneously (i.e. distortions of the signal waveform may be ignored) is not to be confused with the signal being narrowband in the sense that its bandwidth is by far less than that of the carrier, i.e. $\Delta f_s \ll f_c$.

tion coefficient $R(\Delta f)$ of the transmission at frequencies f and $f + \Delta f$. For signals with a bandwidth such that $R(\Delta f) \geqslant 1/e$, the fading may be assumed smooth.

A more exact measure of smoothness should be referred to the permissible distortions of the information transmitted, with the fading considered smooth if the distortions, with the signal waveform and reception method taken into account, do not exceed a predetermined value.

Fading control. The simplest way of suppressing the multiplicative interference is the use of automatic gain control (AGC) in the receiver. Kharkevich has observed [273] that AGC "compensates for the multiplicative interference". But the presence of additive noise $\xi(t)$, with AGC applied, results in fluctuations of the additive interference if the wanted signal has a constant level since the resulting signal is

$$y(t)/|K(t)| = x(t) + \xi(t)/|K(t)| \tag{7.5}$$

the S/N ratio remaining constant.

In analog data transmission, the following methods may be employed to fight multiplicative interference. First, diversity reception constituting a form of the *acquisition* method. This technique consists essentially in forming several replicas of the signal being received, which are differently corrupted by the multiplicative interference, and in combining them thereafter. Second, the *optimal reception* method using spread spectrum signals. Third, the *adaptive reception* drawing on the information about the transmission channel. The adaptation may be effected either in reception or in transmission. The existent methods are illustrated by the diagram of Fig. 7.8.

The use of one or another method of combating the multiplicative interference is conditioned, on the one hand, by reliability requirements and permissible distortions of transmitted information and, on the other hand, by permissible values of parameters and the sophistication of the communication system.

Diversity reception. For years the diversity reception has been the only countermeasure against the fast fading brought about by the multipath structure of the signal at the reception site. Extensive fundamental and experimental evidence has been accumulated, related both to diversity methods and to methods of diversity signals combining [179, 197, 198]. The most common are the space and frequency diversity. In troposcatter links, use is also made of the angle diversity. Any form of diversity requires that the signal fluctuations be statistically independent at the input of diversity receivers. The Rayleigh fluctuations may be thought of as independent if the correlation coefficient for the envelopes RU_1, U_2 is less than 1/e. The improvement gained with diversity reception drops by no

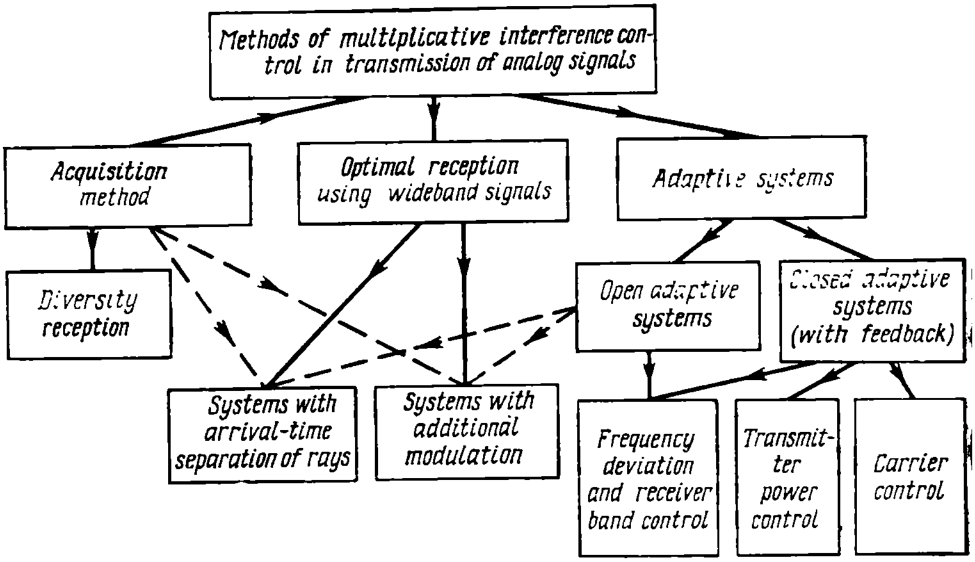

Fig. 7.8. Classification of fading fighting techniques

more than 1.5 dB if the correlation coefficient $R_{U1,U2} \leqslant 0.6$.

After N replicas of the radio signal are available, these should be put to use in a most rational way. In communication systems, the linear combination of diversity signals is adopted, whereby the output signal is

$$y_{\Sigma}(t) = \sum_{i=1}^{N} b_i y_i(t) \tag{7.6}$$

where $y_i(t)$ is the sum of the information-bearing signal and additive noise in the ith diversity path; b_i is the weight factor. The linear combination methods can be appropriately divided into switching and addition methods. In switching, at each given instant of time only one of the coefficients b_i is nonzero. In adding, all the coefficients are nonzero. Of special interest are two addition methods: linear addition in which all the signal replicas received are summed up equally weighted regardless of their level, and optimal addition in which the weight coefficients b_i are automatically controlled so that the S/N ratio at the addition system output be a maximum. In 1965, an addition system was devised for an arbitrary number of EM signals with the use of a deviation subtraction converter (DSC) (Fig. 7.9).

The statistical parameters of N-fold signal reception using various addition methods are discussed, for instance, in [179, 197]. These works, however, treat unmodulated signals. Frequency modulation changes sufficiently the communication system characteristics making it necessary to consider the threshold properties of FM [198].

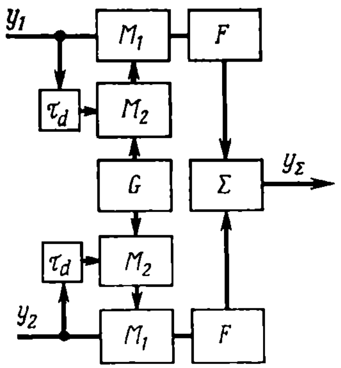

Fig. 7.9. Block diagram of diversity DSC reception; *G*—generator; *M*—mixer, *F*—filter

Moreover, the signal multipath character gives rise to spurious amplitude modulation, crosstalk, and amplitude and phase fluctuations of the signal. In recent years the role of these factors has been considered mainly by Gusyatinsky, Papernov, and Ryskin [91, 92].

Optimal reception methods may be of help in synthesizing the block diagram of an optimal system for combining the diversity FM signals. For large S/N ratios, an optimal demodulator has been synthesized [495], yet its realization appeared to be unfeasible. Proceeding from heuristic assumptions, Van Trees has arrived at a block-diagram of a physically feasible demodulator [495]. The possible options to realize this diagram are given in [42].

Use of signals. The diversity reception may be referred to as a passive way of combating the multiplicative interference, for the transmitted signal does not change until the reception procedure. Transmission of special-purpose spread spectrum signals offers great promise in realization of active methods to control multiplicative interference.

The spread spectrum signal technique bears upon the redundancy introduced to obtain signals with required correlation and spectral properties (see Chap. 5). In digital transmission redundancy is introduced in rather a straightforward manner, e.g., arranging binary 0s and 1s in various orthogonal sequences. In analog transmission, redundancy may not be obviously introduced by such "direct" methods, as in reception it would not be possible to separate the wanted signal from the resultant mixture (even more so if distortions arising in a multipath channel are taken into account). To produce spread spectrum signals a signal obtained by one of the conventional modulation methods rather than the message proper may be used. This signal may be subjected to an additional modulation to obtain correlation or spectral properties required to control the

multipath effect. At the receiving end, methods of spread spectrum signal detection are utilized to produce a mixture of signal and noise which is then demodulated by a conventional technique.

Diversity reception cannot eliminate fading in diversity paths. In spread spectrum signal transmission, interfering can be separated by arrival time in reception. This is accomplished in HF transmission. In the ionoscatter and troposcatter channels the separation of individual paths is not generally possible, because of a lot of scattering irregularities on the propagation path. Nevertheless, here also the representation of the signal as a finite sum of paths is valid. In such a case, the number of paths to be separated is proportional to $2\Delta f \Delta\tau_{max}$, where $\Delta\tau_{max}$ is the maximum difference in delays of the lower and upper paths, and Δf is the bandwidth of the transmitted signal. As the delay in each of detected path is substantially less than $\Delta\tau_{max}$, so are the signal distortions in each receive channel.

Whether or not the paths may be separated is decided by a special spread spectrum signal with a fast-decreasing correlation function [345]. If the peak width of the autocorrelation function is less than $1/2\Delta f$, then the correlation receiver will suppress all the delayed and advanced paths according to the values of the autocorrelation function for the times equal to the delay of the rest of the paths.

Depending on the reception method, either the "strongest" path is singled out or several paths are coherently received and summed up in voltage.

Systems with filtration in time domain. In systems with filtration in time domain, the transmitter generates short pulses with duration $\tau \ll \Delta\tau_{max}$ modulated by a message. There are n receive channels which are in turn gated by the switch for a time T to pass signals with the relative delay under $1/2\Delta f$. Channel outputs are connected via delay lines to a device performing the coherent detector adding. Owing to the delay line, all the signals arrive in the adder simultaneously, the power of the resultant signal, therefore, is equal to the sum of powers of all the components. As in each channel signals are detected with a shorter relative delay, both distortions and depth of fading are reduced. Receive channels gating should be synchronized with the clock frequency of the transmitter, therefore one of the basic components in the receiver is the synchronizer.

The best known systems using an equivalent model of a multipath channel in time domain are the systems of correlation reception, whose concept was put forward in 1958 in [455], and feasibility reported in [398].

A specific embodyment of such a system is the known *Rake* system designed to transmit digital information (for HF channel). In principle, the same concept may be also realized to transmit analog signals, for which purpose the transmitter generates a spread spectrum signal with the correlation function showing a maximum

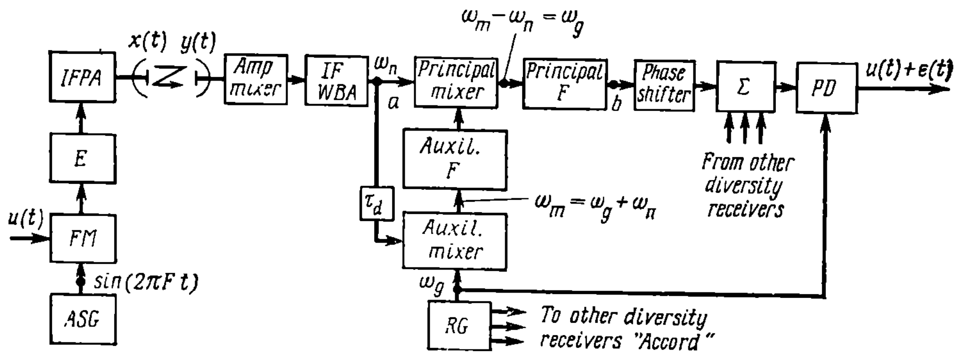

Fig. 7.10. Block diagram of autocorrelation receiver: Accord ASG—auxiliary signal generator; FM—frequency modulator; E—exciter; IFPA—intermediate frequency power amplifier; F—filter; PD—phase detector; RG—reference generator

at zero shift and zeros at a shift by $q\tau_d$ ($q = 1, 2, \ldots$). The receiver should include a number of mutually synchronized correlators.

In transmitting an FM analog signal, the wideband signal may be produced by an additional FM with a sine signal at frequency Ω_{cor}. As the correlation function for such a signal

$$R(\tau) = J_0(2m_{cor} \sin \Omega_{cor}\tau) \tag{7.7}$$

(where m_{cor} is the FM index) considerably differs from the correlation function of the PNS and does not exhibit equidistantly spaced zeros in shifting by $q\tau_d$, the loss in such a system may attain several decibels. It should also be emphasized that in additional modulation other than FM, AM may appear as well, thus leading to a poorer transmitter efficiency.

Systems with filtration in frequency domain. In this case, the transmitter also produces an n-component spread spectrum composite signal. The reception methods may be either similar to those used in time-domain systems or cross-correlated when ancillary signal is synthesized at the receiving site, or autocorrelated when for the demodulation purposes the received mixture of signal and noise is used.

A number of ingeneous receivers of the composite signal in the troposcatter radio channel has been proposed by Gusyatinsky, Plekhanov and co-workers. In 1969, the autocorrelation receiver *Accord* was developed, whose principle is illustrated by the block diagram in Fig. 7.10 [93]. The composite signal, with the components equidistantly spaced in frequency, is formed at the transmitting end by an additional FM carrier that has already been modulated by the message. As a result of this double frequency modulation, the signal

at the transmitter output is

$$u_{tr}(t) = U_{tr} \cos\Big[\omega_0 t + (\Delta\omega_{cor}/\Omega_{cor}) \sin \Omega_{cor} t + \Delta\omega_m \int_0^t u(t_1)\, dt_1\Big] \tag{7.8}$$

or

$$u_{tr}(t) = U_{tr} \sum_{k=-\infty}^{\infty} J_k(m_{cor}) \cos\Big[(\omega_0 + k\Omega_{cor}) t + \Delta\omega_m \int_0^t u(t_1)\, dt_1\Big] \tag{7.9}$$

where $\Delta\omega_{cor}$ is the frequency deviation by an additional sinusoidal wave with frequency Ω_{cor}; $m_{cor} = \Delta\omega_{cor}/\Omega_{cor}$, $\Delta\omega_m$ is the maximal deviation of the frequency by the transmitted message $u(t)$. Having passed the region of multipath propagation, the RF amplifier and the receiver converter, the composite signal at the IF amplifier input takes the form

$$u_s(t) = \sum_{k=-n/2}^{n/2} V_k J_k(m_{cor}) \cos\Big[(\omega_0 + k\Omega_{cor}) t + \Delta\omega_m \int_0^t u(t_1)\, dt_1 + \Phi_k\Big]$$

Here n is the number of components within the receiver passband; $V_k J_k(m_{cor})$ and Φ_k are random amplitudes and phases of signal components, whose fluctuations are due to the multipath propagation of radio waves in the troposphere. At $F_{cor} > \Delta f_0$, these fluctuations can be regarded as statistically independent. To process the composite signal, the receiver uses DSC (see Fig. 7.9). In the main mixer, the deviation of the input signal and the signal delayed by τ_d in the delay line is subtracted. The phase of HF oscillations is

$$\theta(t) = S(t) - S(t-\tau) = \int_0^t u(t)\, dt - \int_0^{t-\tau_d} u(t)\, dt = \int_t^{t-\tau_d} u(t)\, dt \tag{7.10}$$

To achieve an optimum processing of the composite signal, $\tau_d = 1/F_{cor}$ is selected. In this case, the composite signal is delayed in the delay line exactly for a period of the auxiliary sinusoidal signal, and hence, multiplied in the main mixer without the delay, that is squared. As here $\tau_d \ll 1/F_{\mathrm{HF}}$ then from Eq. (7.10) it follows that

$$\theta(t) = \omega_g t + \Delta\omega_m \tau_d\, u(t) - \omega_0 \tau_d \tag{7.11}$$

and further

$$u_{out}(t) = \sum_{k=-n/2}^{n/2} \frac{V_k^2 J_k^2(m_{cor})}{2} \cos \omega_g + \sum_{i,\, j=-n/2\, i \neq j}^{n/2} V_i V_j J_i(m_{cor}) J_j(m_{cor}) \cos[\omega_g + (i+j)\Omega_{cor}]\, t \tag{7.12}$$

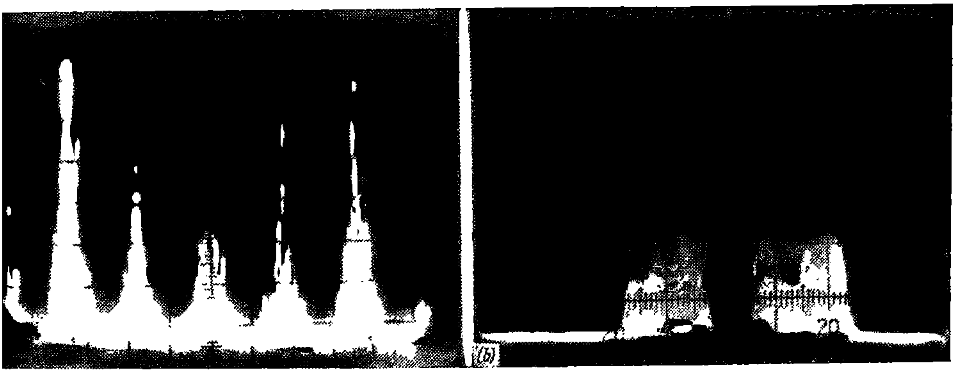

Fig. 7.11. Output spectra of (a) transmitter, (b) receiver filter

The first term in Eq. (7.12) is the central component with the frequency of the reference generator, equal to the sum of squares of all the components of the composite signal. It is separated by the filter of the main mixer. Figure 7.11 shows the photographs of the spectra at the outputs of the transmitter and the filter of the receiver.

Thus, the *Accord* system sums up squares of the components of the composite signal, which provides the optimal addition of signals exceeding the FM threshold. As at the DSC output, the signal phase is independent of the input signal phase and changes but insignificantly with the transmitter frequency drift, the *Accord* system ensures the predetector composition of diversity signals of any multiplicity. This system is employed in the Soviet troposcatter hardware. A further improvement of the autocorrelation receiver *Accord* is the *Saturn* system with signal filtering. The major objective of the development effort has been to eliminate noise multiplication in the main mixer, as this impairs drastically the system performance with small input signals. The block diagram of the *Saturn* receiver is given in Fig. 7.12 [16a]. The feedback circuit consisting of a bandpass filter (BPF) and a comb rejector filter (CRF), a mixer M_2, and a comb filter (CF) serves to produce a reference signal. Here, the FM index is reduced as in a large-index FM the carrier is small and does not lend itself to separation by a filter. To accomplish this, concurrently with the input signal, a signal through the feedback circuit arrives at M_2. If the passage time through the circuit for the signal is $\tau_{fb} = \tau_d$, then the modulation is completely removed. To free from noise, the resultant reference oscillation is filtered by the comb filter. As the circuit of Fig. 7.12 is a regenerative device, to eliminate frequency jumps it is necessary to suppress the generation at frequencies $\omega_{out} \pm \Omega_{cor}$. To this end, a comb rejection filter with rejection at these frequencies is incorporated into the feedback circuit. The

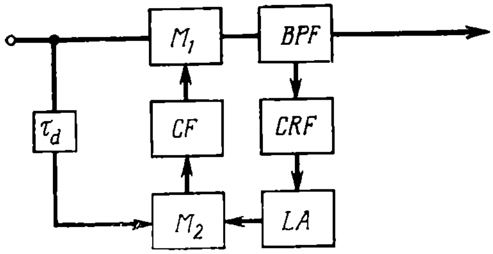

Fig. 7.12. Block diagram of Saturn receiver

limiter prevents the amplitude in feedback circuit from growing, and a linear behavior of multipliers is thereby achieved.

Adaptive systems. There are open and closed adaptive systems. The above systems of correlation reception of the composite signal may be classed with the open adaptive systems processing the signal in reception as here, too, the receiver forms a reference signal determined uniquely by the state of the transmission section.

Closed adaptive systems depend on measurements of the characteristics of the communication channel in reception to work out control signals to be transmitted via the feedback to the transmitter, and it is in the transmitter that appropriate adaptation occurs (at times, the adaptation is performed at both ends of the link). An invariable condition for the development of closed adaptive systems is the availability of a feedback channel. The feedback channel is taken here as the line connecting the receiving end of the system, which evaluates the influence of the propagation medium, with the control device of the transmitting end, which counteracts the changes brought about by the medium thus improving the behavior of the communication system as a whole (feedback systems for discrete data are discussed in [121]).

In order to investigate a transmission channel, both wanted and special-purpose test signals may be utilized. Here the objective is to form such a mixture of wanted and test signals which can be separated with account of the possible interference in the propagation channel. The test signal should adequately reflect the state of the channel and its transmission should not require too much power. Almost all such systems are designed to transmit pulse signals. (With only exception of the system described in [403].)

The feedback may be used to derive information on the signal distortions in the direct channel prior to demodulation ("signal feedback"). Here, information about the level and distortions of the signal is transmitted from receiving to sending end. There may also be a "message feedback", and in that case it connects the demodulator or decoder output with appropriate units at the sending end. In

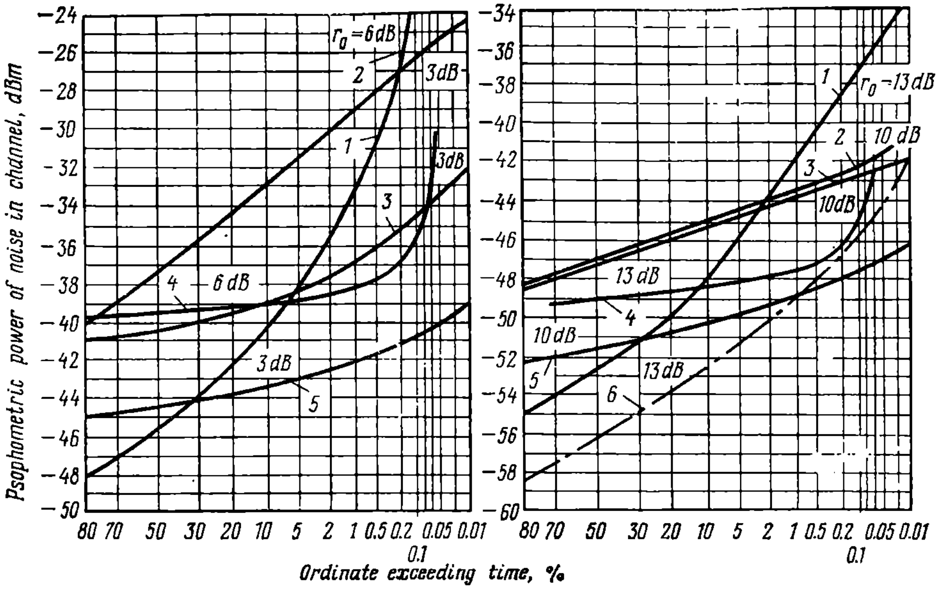

Fig. 7.13. Tradeoff of fading control techniques; $F_c = 256$ kHz; $\Delta f_c = 100$ kHz/channel

the process, not only signals are controlled, but also decisions made by the receiver.

Comparison between multiplicative interference control techniques in troposcatter radio channel. Figure 7.13 depicts the computed noise distributions in a telephone channel at the output of one hop of a radio relay link for the following fading control methods:

optimal addition of diversity signals by using threshold-extending devices (two-input reception), $N = 2, \Delta f = 4$MHz (curve *1*);

autocorrelation two-input reception of a composite signal (*Accord*) (curve *2*, $n = 7$, $N = 2$, $\Delta f = 8$ MHz);

cross-correlation reception of spread spectrum signal (curve *3*, $B = 16$, $\Delta f = 16$ MHz);

use of an adaptive system with automatic control of transmitter power (ATPC) (in two-input reception); making allowance for delay in the control system and limited control range (curve *4*, $N = 2$, $\Delta f = 4$ MHz);

use of an adaptive system with optimal manoeuvring within a wide frequency band (SOMF) without allowance for delay (curve *5*, $F = 16$ MHz, $B = 16$).

In order to compare on an equal basis the ratio of total signal power ($P_s N$) to spectral power density of noise, ($P_n/\Delta F$ MHz) is taken to be similar for all the systems under discussion (in Fig. 7.13*a*, $P_s N/(P_n/\Delta F) = N\Delta F\tau = 32$; and in Fig. 7.13*b*, the ratio is equal

to 160). Here the bands chosen were 16 MHz for the correlation system with wideband signals and SOMF, 2 × 8 MHz for the *Accord* system and 4 MHz for the ATPC and diversity reception systems.

As is readily seen from the figure, the best performance has the adaptive system with feedback—SOMF. (Notably the consideration of delay in feedback circuits may markedly impair the SOMF statistics.) The performance of cross-correlation reception of a wideband signal is 4 to 6 dB worse. The two-input diversity reception even at optimal addition with the use of FM threshold extension devices compares unfavourably with the autocorrelation reception as far as the percentage time determining the system reliability is concerned.

Figure 7.13*b* (curve *6*) presents for comparison the noise distribution at four-input reception with optimal addition and application of threshold-extension techniques. It is seen that for small percentage time even at $N = 4$ the diversity reception is inferior to SOMF and approaches the two-input reception with the use of the *Accord* system.

7.4. Electromagnetic Compatibility of Radiocommunication Systems

Figures of merit of radiocommunication systems. The development of radio communication and radio systems, and ever growing crowding in usable frequency range give rise to interference signals, i.e. unintentional radio interference. Not only the number of radio emissions is increased, but also the power of transmitters, and the sensitivity of receivers. Wideband modulation methods are bringing into extensive use. The requirements for simultaneous operation of many radio communication systems under these conditions have called for fundamental and practical investigations into the electromagnetic compatibility (EMC) of radio electronic facilities. The aim of these investigations is the development and realization of measures to reduce the impact of unintentional radio interference on major parameters of radio facilities and systems.

We now consider some issues of EMC of line-of-sight radio relay lines and satellite communication systems. These modern facilities feature immense capacity and, according to Radio Regulations [225], operate within the common frequency bands.

In analog communication systems, when telephone messages are transmitted, the major figure of merit defining the permissible impact of interfering radio signals is the ratio of the measuring signal power $P^*_{s,out}$ to the power of noise (disturbing signals) $P^*_{d,out}$ at the channel output

$$Q^*_{out} = P^*_{s,out}/P^*_{d,out} \tag{7.13}$$

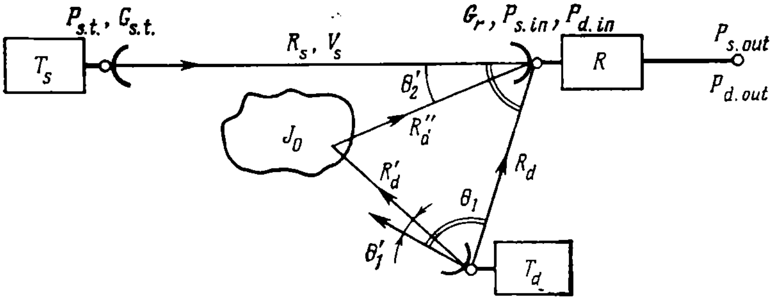

Fig. 7.14. Signal-to-interference relationships

For television broadcasts, the ratio of powers of information-bearing and disturbing radio signals, obtained at the receiver input from measurements of picture quality [157], is normally also specified.

With digital communication systems, the basic parameter is the permissible error probability p_e^*.

It stands to reason that the parameters of realized systems, $P_{d,out}$, Q_{out}, p_e, should satisfy the conditions $P_{d,out} \leqslant P_{d,out}^*$, $Q_{out} \geqslant Q_{out}^*$, $p_e \leqslant p_e^*$. The quantities Q_{out} and p_e are best represented as the products

$$Q_{out} = Q_{in}\varkappa \tag{7.14}$$

$$p_e = Q_{in}s \tag{7.14a}$$

where $Q_{in} = P_{s,in}/P_{d,in}$ is the power ratio of information-bearing and disturbing signals at the receiver input, which is dependent on the powers of message signal and disturbing transmitters, $P_{s,t}$ and $P_{d,t}$, antenna gain with regard to their orientation $G_{s,t}$, $G_{d,t}$, G_r, distances from the transmitting antennas to the receiver, R_s and R_d, and attenuation factors V_s and V_d. Thus, (see Fig. 7.14)

$$Q_{in} = f(P_{s,t}, P_{d,t}, R_s, R_d, G_{s,t}, G_{d,t}, G_r, V_s, V_d) \tag{7.15}$$

Note that in rationally constructed communication systems, Q_{in} may be regarded as independent of the type of modulation and structure of the message transmitted. Owing to the time variation in V_s and V_d and in orientation of antennas, Q_{in} also varies with time. The quantities $\varkappa$ and s are determined by the ratios: $P_{s,in}/P_{d,in}$, $P_{s,in}/P_{n,in}$ (here $P_{n,in}$ is the power of thermal noise as referred to the receiver input), by frequency stability by modulation type, by spectra of information-bearing and disturbing signals and their overlap, by the structure of transmitted messages, by responses of individual units and all the receiving system. These quantities are also time variant. However, to simplify the calculations, it would pay to determine $\varkappa$ and s for the predetermined cases in

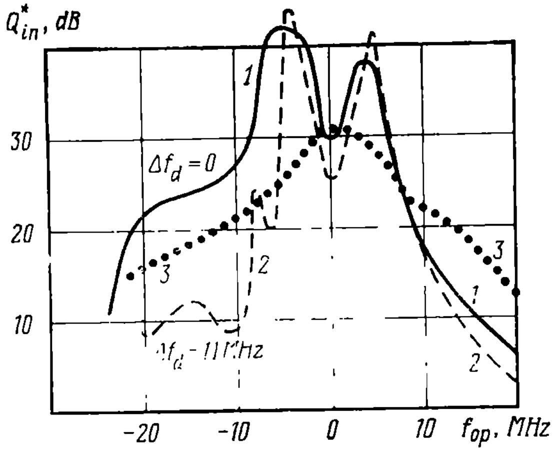

Fig. 7.15. Signal-to-interference ratios for (1) disturbing signal not modulated; (2) frequency modulation with peak deviation 11 MHz; information-bearing signal modulated by colour test signal; (3) information-bearing and disturbing signals modulated by black-and-white TV message

which the ratios $P_{s,in}/P_{d,in}$ and $P_{s,in}/P_{n,in}$ exceed the specified percentage times, and a channel loading that would be the most unfavorable in interference sense. In tradeoffs of various systems, reception methods and signal processing techniques, the quantities $\varkappa$ and s may be taken to be a measure of the electromagnetic compatibility.

When dealing with communication links consisting of n intervals that contain $n - 1$ repeaters and with no correcting means,

$$Q^{-1}_{out} = \sum_{i=1}^{n} (\varkappa_i Q_{in,\,i})^{-1} \tag{7.16}$$

$$p_e = \sum_{i=1}^{n} s_i Q_{in,\,i} \tag{7.17}$$

assuming that the terms in Eq. (7.17) are so small that for real values of n the sum is much less than unity and meets the specifications.

With independent disturbing signals in one or several intervals, calculations should take into account all these signals.

It should be noted that for such large-capacity systems as radio relay lines and satellite communication networks, the Radio Regulations and CCIR recommend the conditions and modes for the station to meet the condition $P_{d,\,in} \ll P_{s,\,in}$ [192, 225]. Therefore, in those systems one may neglect such phenomena as the masking of the information-bearing signal by the disturbing, and existence of extraband and intermodulated reception channels. This in large measure facilitates the determination of values of $\varkappa$, s, $P_{s,out}$ and p_e.

Determination of function Q_{in}. It follows from Eq. (7.15) that the determination of Q_{in} requires a knowledge of powers of input information-bearing signals $P_{s,in\ 0}$ and input disturbing signals $P_{d,in\ 0}$ for propagation in free space. These powers are derived from the relationships

$$P_{s,\ in\ 0}=\frac{P_{s,\ t}G_{s,\ t}G_r\eta_{s,\ t,\ f}\eta_{r,\ f}\lambda_s^2}{16\pi^2R_s^2}$$

$$P_{d,\ in\ 0}=\frac{P_{d,\ t}G_{d,\ t}G_r\eta_{d,\ t,\ f}\eta_{r,\ f}\lambda_d^2}{16\pi^2R_d^2}$$

where η and λ are the feeder efficiency and wavelength, respectively; the sense of the other quantities is clear from Fig. 7.14. The values of G are determined from averaged data obtained for a large number of antennas implemented in various countries [116, 119, 192, 225].

It follows from Eq. (7.15) that

$$Q_{in}=q_0V_s^2/V_d^2 \tag{7.18}$$

where $q_0 = P_{s,\ in\ 0}/P_{d,\ in\ 0}$.

Calculations from Eq. (7.18) should be performed not for the whole time period, but for the time that is defined by the given value of Q^*_{out} or p^*_e. The analytical expressions for attenuation factors V_s and V_d for various conditions are to be found in [113, 120]. The determination of V_s at frequencies higher than 8-10 GHz should take into account the attenuation V_ϱ in precipitation, and the determination of $P_{d,in}$ the power $P_{d,\ in,pr}$ of signals reflected by precipitation (see Fig. 7.14).

It has been noted earlier that in transmitting television broadcasts in an analog form, the desired value of Q_{in} is taken as is determined from the subjective statistical test. Figure 7.15 presents the curves for three cases [192].

Impact of disturbing signal on FM receiver. Now we carry out an analysis for $P_{d,in} \ll P_{s,in}$ and $P_{n,in} \ll P_{s,\ in}$. In that case, the impact of disturbing signal and thermal noise on the frequency detector may be approached separately, and the power spectrum at the FD output and the noise in channels may be obtained by summing up, respectively, power spectra and noise due to disturbing signal with thermal noise. The instantaneous values of voltages of the information-bearing and disturbing FM signals at the receiver input are

$$u_{s,\ in}=U_{s,\ in}\cos\alpha \tag{7.19}$$

$$u_{d,\ in}=U_{d,\ in}\cos\beta \tag{7.20}$$

$$\alpha=\omega_st+\Delta\omega_{\mathrm{rms},\ s}\int_{-\infty}^{t}\xi_s(t)\,dt \tag{7.21}$$

$$\beta=\omega_dt+\Delta\omega_{\mathrm{rms},\ d}\int_{-\infty}^{t}\xi_d(t)\,dt \tag{7.22}$$

where $\xi_s(t)$ and $\xi_d(t)$ are the modulating random processes; $\Delta\omega_{rms,s}$ and $\Delta\omega_{rms,d}$ are the rms values of deviation. Setting $k = U_{d,in}/U_{s,in}$, $\omega_{dif} = |\omega_d - \omega_s|$, $\psi = \beta - \alpha$ we arrive at

$$U_{\Sigma in} = U_{s,in}\sqrt{1+k^2+2k\cos\psi} \tag{7.23}$$

The resultant signal at the input will be

$$u_{\Sigma in} = U_{\Sigma in}\cos(\alpha+\vartheta) \tag{7.24}$$

where

$$\vartheta = \operatorname{arctg}[k\sin\psi/(1+k\cos\psi)] \tag{7.25}$$

If $k \ll 1$, then

$$\vartheta \approx \operatorname{arctg}(k\sin\psi) \tag{7.26}$$

At the output of an ideal frequency detector with the slope S_{FD}, when Eqs. (7.24) and (7.21) are considered, we have

$$u_{out} = S_{FD}[\Delta\omega_{rms,s}\xi_s(t) + d\vartheta/dt] \tag{7.27}$$

With a correctly established level diagram to feed a measuring signal which sets up a specified rms value of the deviation $\Delta\omega_k$, the rms voltage at the FD output is

$$U_{s,out,rms} = S_{FD}\Delta\omega_k \tag{7.28}$$

The first term in brackets of Eq. (7.27) is the undistorted part of the message, and the second the interference. It follows from Eqs. (7.28) and (7.27) that

$$U_{d,out} = S_{FD}\frac{d\vartheta}{dt} = \frac{U_{s,out,rms}}{\Delta\omega_k}\frac{d\vartheta}{dt} \tag{7.29}$$

To obtain the power spectrum of the interference at the FD output, we use the Wiener-Khinchin theorem [162]. After some manipulations we write the expression for power spectrum at the FD output in the resistance R_{out} [118]

$$F_{d,out}(f) = \frac{P_{s,out}R_{out}}{2}\left(\frac{f}{\Delta f_k}\right)^2 \sum_{n-1}^{\infty}\frac{k^{2n}}{n^2}J_n(f) \tag{7.30}$$

where

$$J_n(f) = \int_{-\infty}^{\infty} F_{dn}(z)\,[F_{sn}(nf_{op}-f-z) + F_{sn}(nf_{op}+f-z)]\,dz, \quad 2\pi f = \omega \tag{7.31}$$

F_{sn} and F_{dn} are the power spectra of the processes $\cos\left[n\Delta\omega_{rms,s}\int_{-\infty}^{t}\xi_s(t)\,dt\right]$ and $\cos\left[n\Delta\omega_{rms,d}\int_{-\infty}^{t}\xi_d(t)\,dt\right]$, respectively. We find, by Eq. (7.30), the power of interference in a te-

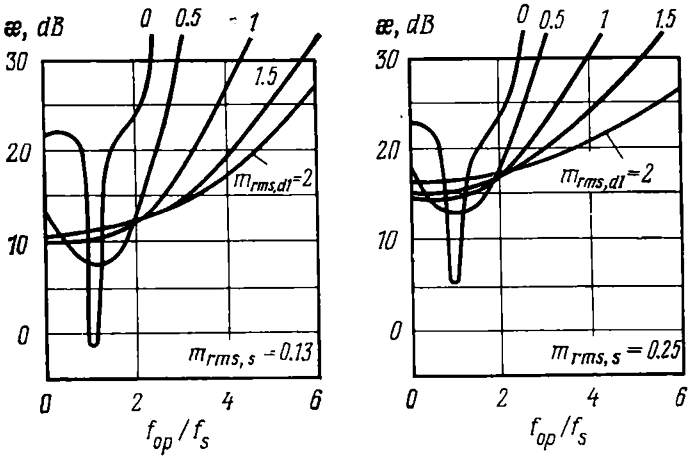

Fig. 7.16. EMC criterion for one-hop link

lephone channel with bound frequencies f_{1k} and f_{2k}:

$$P_{d,out} = \frac{1}{R_{out}} \int_{f_{1k}}^{f_{2k}} F_{d,out}(f)\, df \tag{7.32}$$

Substituting Eq. (7.32) into Eqs. (7.14) and (7.15) we can derive the EMC criterion for a one-section link

$$\varkappa = k^2 Q_{out} = P_{s,\,out} P_{d,\,in} / P_{d,\,out} P_{s,\,in} \tag{7.33}$$

The values of $\varkappa$ in decibels for the upper telephone channel with central frequency f_c and pre-emphasis of the spectrum of the multichannel telephone communication recommended by the CCIR are given in Fig. 7.16, where

$$m_{ef,\,s} = \Delta f_{\mathrm{rms},\,s} / f_c \quad m_{ef,\,d} = \Delta f_{\mathrm{rms},\,d} / f_c$$

Note that the expressions for the power spectrum of the useful and disturbing signals at small and large modulation indexes have been derived in [162], and for $m \approx 1$ in [117].

A scrutiny of Fig. 7.16 suggests that a reduction in parameter $\varkappa$ occurs with decreasing $m_{ef,\,s}$ and $m_{ef,\,d}$. Therefore, in changing the loading of the communication system and in redistributing the loading from lower channels to the upper ones, stronger interference will result in telephone channels.

Especially strong interference in telephone channels will be observed without modulation of the information-bearing and disturbing signals, i.e. at $\Delta f_{\mathrm{rms},\,s} = \Delta f_{\mathrm{rms},\,d} = 0$. In that case, when $f_{1k} \leqslant f_{dif} \leqslant f_{2k}$, where $f_{dif} = |f_{0d} - f_{0s}|$ and the rms deviation cau-

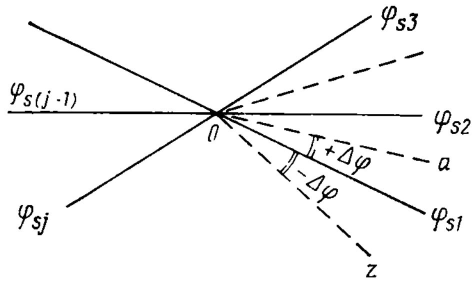

Fig. 7.17. Phase angles for various signals

sed by the measuring signal is Δf_k, we have [119]

$$P_{d,\,out} = P_{s,\,out} k^2 f_k^2 / 2\Delta f_k \tag{7.34}$$

$$\varkappa = 2\,(\Delta f_k / f_k)^2 \tag{7.35}$$

It should be emphasized that owing to instability of oscillations emitted by the transmitters of the information-bearing and disturbing signals, the quantity f_{dif} will change, i.e. the interference will migrate from one telephone channel to another gaining in strength when coming to the upper telephone channels.

The investigation of the effect on the FM receiver caused by disturbing signal and noise, whose powers are compatible with that of the information-bearing signal, is the subject of a number of works, e.g. [76, 77, 124].

Impact of disturbing signals on PSK receiver. Let us consider the coherent receiver as exposed to the action of a PSK information-bearing signal (deterministic)

$$u_s(t) = \cos[\omega_s t + \varphi_s(t)] \tag{7.36}$$

concurrently with the disturbing signals and noise. The phase $\varphi_s(t)$ may take on fixed values $\varphi_{s1}, \varphi_{s2}, \ldots, \varphi_{sM}$ ranging from 0 to 2π. It is often assumed that

$$\varphi_{sj} - \varphi_{s(j-1)} = 2\pi/2^M \tag{7.37}$$

Indicated in Fig. 7.17 are several sequential values of the phase angle for various symbols to be transmitted. The receiver performs the inverse transformation. We will regard the phase detector (PD) as ideal with zero values of thresholds defining the bounds of symbol variation shown in Fig. 7.17 by the dashed lines *oa* and *oz*. If under the action of noise or disturbing signals the composite vector appears beyond the boundary lines, a false symbol will be detected. We find, by Eq. (7.37), the angle $\Delta\varphi$ beyond which an error arises

$$\Delta\varphi = \pi/2^M \tag{7.38}$$

And the disturbing signal we represent as

$$u_d(t) = k\cos[\omega_d t + \varphi_d(t) + \varphi_0(t)] \tag{7.39}$$

where $\varphi_d(t)$ determines the phase modulation of the signal due to the message being transmitted, and $\varphi_0(t)$ is the change in the phase angle caused by the instability of performance of the disturbing transmitter, whose probability distribution is uniform in the interval $(0, 2\pi)$.

We represent the thermal noise as two independent normally distributed quadrature components

$$n(t) = u(t)\cos\omega_s t + v(t)\sin\omega_s t \tag{7.40}$$

In coherent reception, one multiplier input is exposed to the resultant voltage

$$u_\Sigma(t) = u_s(t) + u_d(t) + n(t) \tag{7.41}$$

and the other to

$$u_0(t) = 2U_0\cos\omega_s t \tag{7.42}$$

At the filter output we obtain

$$U_\Sigma = U_0(X_\Sigma^2 + Y_\Sigma^2)^{1/2} \tag{7.43}$$

where

$$X_\Sigma = 1 + k\cos[\omega_{op}t + \gamma(t)] + u(t) \tag{7.44}$$

$$Y_\Sigma = k\sin[\omega_{op}t + \gamma(t)] + v(t) \tag{7.45}$$

$$\gamma(t) = \varphi_d(t) + \varphi_0(t) \tag{7.46}$$

The error arises if

$$|\alpha_\Sigma| \geqslant \Delta\varphi \tag{7.47}$$

where α_Σ is the phase of the composite vector.

If the prior probabilities of any symbol transmission are equal, then for the just-considered signal and thermal noise models it is not difficult to find the probability density for the random quantity α_Σ and to determine the error probability [119]:

$$p_e = \mathrm{P}\{|\alpha_\Sigma| \geqslant \Delta\varphi\} \tag{7.48}$$

Figure 7.18 gives the results of computations of p_e for various values of $P_{s,\,in}/P_{d,\,in} = k^{-2}$ and $P_{s,\,in}/P_{n,\,in} = 1/2\sigma^2$. From the curves, the value of s may be obtained which is a measure of EMC for digital systems. The curves for other values of M are available in [119].

EMC improvement. Measures and recommendations pertaining to EMC improvement in communication systems, i.e. to abatement of the action of unintentional radio interference on the communi-

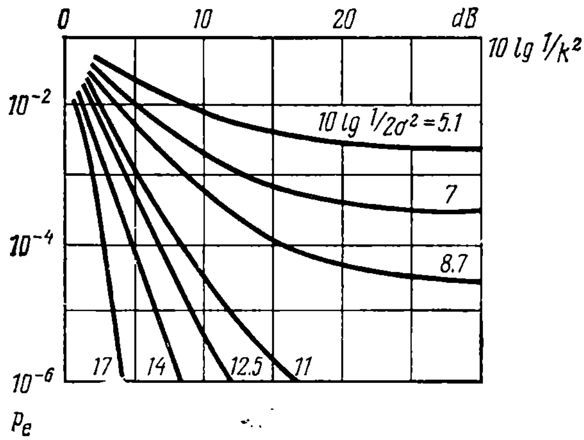

Fig. 7.18. Error probability curves

cation system performance, may be divided into several groups.

The first group includes the measures associated with the rational use of frequency bands by all forms of radio services (definition of services permitting simultaneous operation in common frequency bands, definition of recommendations as to the frequency stability, amount of side lobes, power constraints, etc.). These measures are described in [13, 14, 119, 192, 225].

The second group involves the measures that improve EMC through perfection of individual elements of communication systems, i.e. those measures which are aimed at reduction of side lobes of antenna directional pattern, reduction in the number of side reception channels, improvement of receiver selectivity, etc. [3, 11, 88, 102].

The third group relates to the recommendations as to the signal variance and separation of stations in space [116, 119].

7.5. Optical Communication Systems

Introductory remarks. After the laser was invented in 1961, intensive attempts started to adapt it for the transmission of intelligence. But the characteristics of the atmosphere that was used as a transmission medium in the earliest systems turned out to be unsuitable for highly reliable systems. The attenuation of the laser radiation in rain, fog, and snow is as high as 120 dB/km. To insulate the channel from the environment, light guides have been devised which are essentially metal pipes with a periodic correction of beam divergence and direction. This correction used to be performed with a system of lenses and mirrors. Light guides with a discrete correction were found to be rather costly, requiring painstaking aligning and sophisticated devices for automatic correction [43].

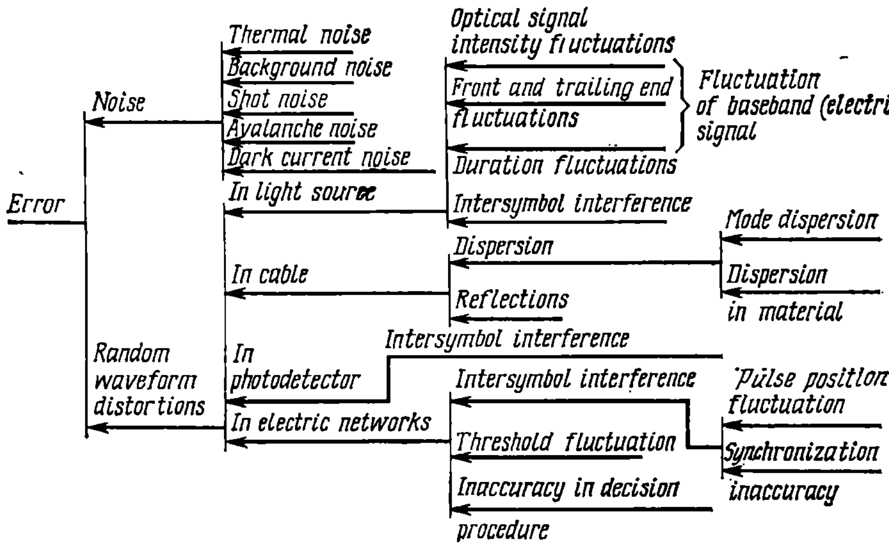

Fig. 7.19. Distortions and noise in optical cable communication line

Extensive efforts were initiated to develop optical communication systems in 1972-73 when low-attenuation glass fibers were produced to form the basis for the creation of optical cables. At present, in many countries, optical cable systems are being worked out (city-exchange, intercity trunks, district networks) with a transmission rate from 2 to 400 Mbit/s.

Optical cable systems offer a number of advantages. They are wideband, low-attenuation (for the best fibers less than 0.5 dB/km), small-sized and light-weight systems; they do without expensive and scarce metals and are distinguished for their immunity to electromagnetic interference, no risk of sparking and ingition in short-circuiting, amenability to light isolation of one fiber from another.

As optical cables compare favourably with HF symmetrical and coaxial cables, the line channel with the optical cable seems to find application in all the sections of the communication system, from local exchange to long-haul links, in distributed computer systems, and also in special-purpose systems.

Optical vs. conventional systems. The scheme of noise and distortions responsible for errors in transmission over the optical cable system is represented in Fig. 7.19. Central to the system are the properties of optical cable [195]. The construction of an optimal receiver makes only allowance for the noise associated with the processes of photodetecting and amplifying. Generally, the noise is modeled by the filtered Poisson process.

The structure of the optimal receiver is obtained either using the likelihood functional or minimizing the Chernov bounds for the error probability. In the general case, the optimal receiver registers the instants when photoelectrons appear. The practical realization of optimal receivers requires rather sophisticated high-speed computers [366, 391, 437]. In actual practice, simpler receivers are applied, which are not much inferior to optimal. To receive pulses after the fiber line, use is made of an integrator with reset, a filter matched with the received waveform, and a low-pass RC-filter [196, 451].

Intermediate repeaters (regeneration amplifiers) account for a substantial percentage of the total cost of data transmission systems, and therefore it pays to maximize their separation. In optical cable systems, this separation is limited, for one thing, by thermal noise and photodetecting noise (shot, avalanche, dark current, background), and for the other, by intersymbol interference connected with waveform changing and broadening during propagation along the optical cable, the interference growing fast with increasing the propagation speed and distance.

In order to diminish the intersymbol interference in the receiver after the photodetector but before the decision making device, it would be advisable (as with symmetrical and coaxial cables) to compensate for the pulse shape and duration using equalizer networks. These latter may be linear and digital.

The major condition for effective equalizing is the channel linearity. Though the photodetector has a square-law behavior, in certain practically important cases overlapping pulses may be separated at the receiving end. Such cases include: a laser operating with a single spatial mode, and a multimode light guide; a light-emitting diode and a multimode light guide; a laser operating with mode synchronization and a multimode light guide with inhomogeneities; an incoherent source and a multimode light-emitting diode with inhomogeneities.

The equalizer performance is the result of the compromise between a lower intersymbol interference and a higher noise level (associated with photodetecting and amplifying) in minimal level of required light power [80].

The hop length l is obtainable from the condition that $W_0 - W(l) = \gamma l$, where W_0 in dB m is the light power at the cable input; $W(l)$ in dB m is the power required at the photodetector as a function of l, as the duration τ_0 of received pulses is dependent on l; γ in dB/km is the attenuation in the cable. The duration of pulses τ at the equalizer output is chosen from the condition that the intersymbol interference is small.

There are two types of interference (intersymbol interference and noise of detecting and amplifying) with the result that the optical

cable line has an optimum in pulse ratio. An increase in the pulse ratio leads, for one thing, to an increase in the pulse energy, and on the other thing, to an increase in the light power required at the photodetector to maintain the error probability below a specified level. The increase in required power is due to dispersive behavior of the optical cable; pulses broaden and overlap in propagating (intersymbol interference). With a higher pulse ratio, the overlapping of pulses grows and a higher degree of equalizing is required (boosting of the frequency response). In the process, the effect of noise associated with the photodetecting process and the first stage of postdetector amplification grows, and hence the required light power increases. Thus, an optimal signaling interval pulse ratio exists which yields the largest length of the hop.

With heavy interference, the linear filters become inefficient. Among digital filters those based on the Viterby algorithm deserve mention; their behavior is essentially similar to that of the maximum likelihood receiver. If the pulses are spread to the neighbouring interval only, the algorithm is realized by a relatively simple device.

Multiplexing of optical channels. The discrete character of signals and noise and the square-law characteristics of the photodetector leads to the fact that the quality of reception of pulsed signals is dependent not only on the energy of pulses, but also on their waveform and duration. The best pulse from this viewpoint is the maximally short. It is desirable to transmit binary sequences by a pulse and pause, i.e. to concentrate the energy in a maximally short interval.

In the optical range the following multiplex methods are utilized.

Optical frequency division multiplex. Radiation is introduced into a fiber from several sources operating in various frequency ranges lying within the transparency band of the fiber; the multiplexing is accomplished by optical filters.

Optical time division multiplex. This technique ensures fuller utilization of the capacity of the optical cable with low-frequency terminal equipment. The light source (laser diode) operates under short-pulse conditions (mode synchronization) with a low pulse ratio. The laser beam is divided with the aid of a set of semitransparent mirrors into several beams each of which is modulated by individual signals. The modulation consists in pulsed gating of a beam. The light beams thus modulated are again combined into one, but with an appropriate time delay. The delay is so selected that the pulses of the second, third, etc. channels are arranged after the pulse of the first channel within its pulse spacing.

For channel division in such a pulse stream, sinusoidal control signals are used (applied, for instance, to turn the polarization plane) with frequency $1/2T$, where T is the pulse repetition period in

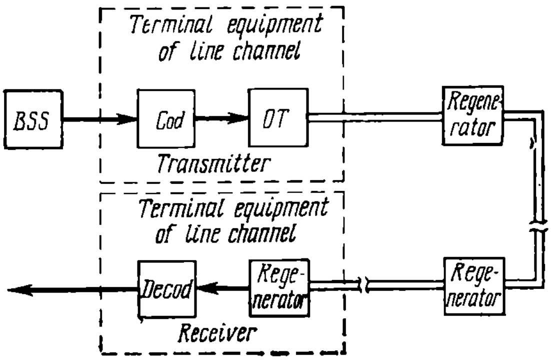

Fig 7.20. Block diagram of optical cable digital data transmission system: BSS—binary signal source, OT—optical transmitte-

the multiplexed sequence. One of the advantages of this method as regards the system requirements is the fact that oscillations controlling the modulator have essentially a zero band.

Another way of time division relies on the space scanning of the multiplexed beam. If the scanning period corresponds to the pulse spacing in the initial sequence, then at certain instants of time the beam will pass through the points in space which correspond to the time division channels. By placing gating detectors at these points, individual channels may be restored.

Space division multiplex. The cable is composed of fibers covered with a nontransparent layer. The quality light-isolation of fibers in the optical cable is readily achieved (in commercial cables the fiber junction attenuation is 80 dB/km). Small fiber cross-section enables in one cable a great many fibers to be combined. In this method, the problem of channel integration and separation is solved easily, therefore current systems adopt mainly this method.

In each of the above-mentioned methods of optical multiplex additional multiplexing can be achieved in electric signal, either by FDM (using subcarriers) or TDM techniques. The baseband signal, in turn, may be a group signal corresponding to a multichannel communication. Current systems use mainly time division multiplex. To transmit analog information (television, facsimile), use is made of the interval pulse-position modulation (IPPM) [165]. With PPM, information is carried by the interval between the clock pulse and signal pulse, and with IPPM, between neighbouring signal pulses. The transmission of digital information uses almost exclusively the pulse-code modulation distinguished by its improved noise immunity and low accumulation of noise.

Signal selection. Figures 7.20 and 7.21 show block diagrams of an optical digital system and a linear regenerator.

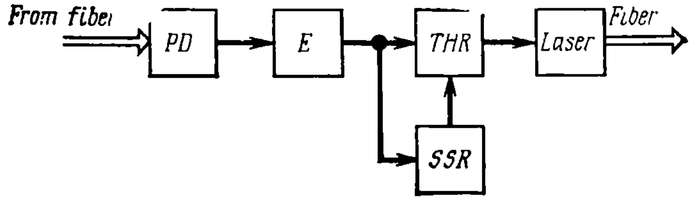

Fig. 7.21. Block diagram of linear regenerator: PD—photodetector; E—equilizer; THR—threshold device SSR—sync signal recovery

When choosing a linear signal (code), two groups of factors are to be taken into account: first, bit error rate of the code that defines the minimum mean power of the signal in the link, required to ensure a given quality of reception in the presence of nonremovable noise (photodetecting, avalanche multiplication, and thermal noises), provided the equipment operates ideally (i.e. ideal synchronization, absence of threshold fluctuations, shifts, etc.); second, physical realization of the code, including the low transmission rate in the link, simplicity of the encoder and decoder (i.e. converters of the terminal equipment to the baseband signals of the link, and vice versa), simplicity of synchronization; minimal content of low-frequency components in the code, possibility of error detection.

As far as the error rate is concerned, the PCM signal is optimal for the optical range: a 1 is transmitted by a pulse, a 0 by a pause. In order that the requirements of the second group may be met, it is necessary to introduce redundancy. This is achievable in two ways, by raising the transmission rate in the link, or by increasing the number of levels. The second approach presents the following problems in optical cable systems. On the one hand, the nonlinear modulation behavior of the laser diode necessitates nonuniform separation of the signal levels and thresholds. On the other hand, intersymbol interference results in a much faster growth of the required power than is the case with two-level systems. Therefore, we will concern ourselves with two-level codes only.

Depicted in Fig. 7.22 are (*a*) single current binary code, (*b*) three-level code widely used in systems with coaxial and symmetrical cables; (*c*), (*d*), and (*e*) double current binary signals with doubled transmission rate ($1B2B$). In the general case the block codes $mBnB$ may be utilized, in which the required increase of rate (n/m) is less than twofold (Fig. 7.23).

In order to estimate and compare block codes which transform $(b_1, \ldots, b_m)$, a binary word of length m, to a code word $(a_1, \ldots, a_n)$ formed by n binary symbols at $n > m$ (codes ${}_mB_nB$), the following parameters are used:

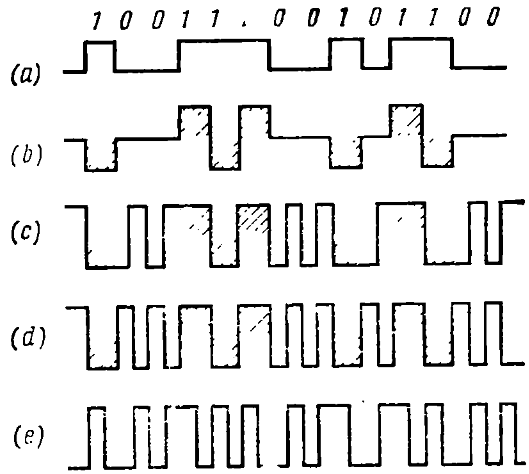

Fig. 7.22. Codes (*a*) binary; (*b*) three-level; (*c*), (*d*), (*f*) two-level

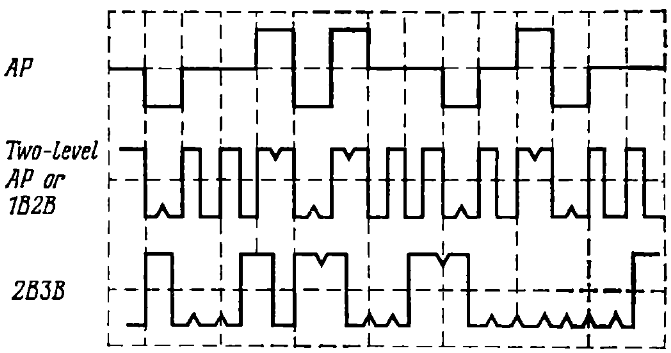

Fig. 7.23. Code transformation

mean symbol length $\langle Q_k \rangle$ that should be minimized in order that the optical power and shot noise of the photodetector may be reduced; if the number of 1s is equal to that of 0s, then $\langle Q_k \rangle = 1/2$,

number of states of encoder, S_k, determining its complexity, and also the complexity of the control circuit used in the decoder to restore synchronism; this parameter coincides with the number of possible values taken on by the current digital sum at the end of each code word;

number of values of S_m which the current digital sum may get at any instant of time; its magnitude dictates the complexity of the error check circuitry used in the linear regenerator;

maximum number of 0s and 1s in succession (l_{0max} and l_{1max})

permitted for the signal; these parameters are best selected minimal to simplify the channel of clock recovery;

continuous part of the power spectrum Δ_1 and Δ_2 (in per cent) lying in the range (0 to 0.03) P or (0 to 0.1) P, respectively, where T is the pulse interval;

ratio of the transmission rate over the link (link symbols) to the binary signal rate at the source output (n/m), which characterizes an increase in the overall transmission rate;

disparity (excess of the number of 1s over the number of 0s); the lower the disparity, the simpler the synchronization channel and the lower the content of low-frequency components in the spectrum (D is the disparity of one combination, D' is the accumulated disparity);

power losses associated with the production of the optimal waveform of the optical signal.

At the receiving end, a device is installed to check decoder operation and restore block synchronism. This device (should the error rate exceed a certain threshold) shifts the signal of block synchroni-

TABLE 7.2

Reception	Type of signals	Additional optical power, dB: wideband fiber	Additional optical power, dB: narrowband fiber	Synchronization: maximum number of symbols in continuous succession	Synchronization: maximum-to-minimum ratio for number of transitions per block	Relative bandwidth	Error check	Dependence of threshold on signal level fluctuations	Encoder complexity (number of ICs)
Symbol-by-symbol	Single current binary AP	0	0	××)	××)	1.0	Impossible	Strong	0 (5)
		7.2	3	×)	×)	1.0	Possible	Strong	4
	Double current binary AP	3.5	$\geqslant 3.5$	2	2 : 1	2.0	—»—	Weak	4
	2B3B with bias	2	$\geqslant 2$	7	2 : 1	1.5	—»—	—»—	11
	2B3B without bias	2	$\geqslant 2$	5	3 : 1	1.5	Impossible	—»—	16
	3B4B without bias	1.5	$\geqslant 1.5$	5	4 : 1	1.3	Possible	—»—	22
Delayed sequence	Double current binary AP	~ 5	$\leqslant 2$	×)	×)	1.0	—»—	Strong	4
	Modified duobinary	~ 5	$\leqslant 2$	××)	××)	0.5-1.0	—»—	—»—	4
	Modified duobinary II cl.	~ 5	$\leqslant 2$	××)	××)	1.0	—»—	—»—	4

zation in succession by one symbol until the value of the current digital sum of the block appears within given limits.

Apart from the structure, signals may differ in the way of reception (detection). In addition to the symbol-by-symbol detection, procedures are of interest that deal with direct and delayed sequences.

TABLE 7.3

Code	n/m	l	Disparity bounds		Code	n/m	l	Disparity bounds	
			D	D'				D	D^1
7B10B	1.43	8	8	±4	5B6B	1.20	6	±2	±4
3B4B	1.33	4	±2	±3	7B8B	1.14	9	±4	±7
6B8B	1.33	6	0	±3	9B10B	1.11	10	±4	±8

Tables 7.2 and 7.3 summarize some of the characteristics of the considered signals, and Table 7.4 provides the characteristics of certain codes $_mB_nB$ [489].

TABLE 7.4

Code	n/m	$\langle\theta_k\rangle$	S_k	S_m	l_0 max	l_1 max	Δ_1, %	Δ_2, %
5B6B	1.2	0.5	2	9	6	6	0.50	5.2
5B7B	1.4	0.42	1	17	6	4	0.20	4.0
2B3B	1.5	0.33	2	8	7	3	0.42	4.3
3B4B	1.33	0.5	2	5	4	4	0.25	4.4
CM1	2	0.5	2	4	3	3	0.11	3.7
1B2B	2	0.5	2	3	2	2	0.03	0.7

Chapter

8

Communication Networks

8.1. Auxiliary Information in Multiaccess Systems and Computer Networks

In any information transmission system, including such a complex one as a computer network, nondetermined factors should be taken into account. Their adverse effect may be offset by using an auxiliary information system whose task reduces to an estimation of varying (a priori unknown) factors affecting the functioning of the main system. The first three sections of this chapter describe the role of auxiliary information in complex systems with many users. The auxiliary information permits shaping more effectively in time the flow of signals introduced into the channel and determine more rationally the channel sequence (route) over which signals are communicated in the network. The second function might also be called the shaping of signal flow in space. Clearly, in the general case the combined influence of auxiliary information on forming both time and space structure of signal flows should be considered. The problem being exceedingly difficult, we consider separately the influence of auxiliary information on the forming of flows in time and their shaping in space. The first problem will be taken in Section 8.2 for a relatively simple system involving all the major problems of the signal flow shaping in time, but not in space, namely remote multiaccess systems in which individual transmitters have access to the common channel through local trunks only. Section 8.3 considers the problem of signal shaping in space, that is reduced to the derivation of a routing technique.

In our reasoning we will proceed from the fact that the task of a system is to provide communication between the elements of the distributed computer network (terminals, processors, etc.). Information (message) in such a system has a specific block-hierarchical structure (see [473]), i.e. it is essentially a sequence of nonuniformly distributed blocks separated by time pauses. These blocks, in turn, are also sequences of smaller subblocks separated by smaller pauses,

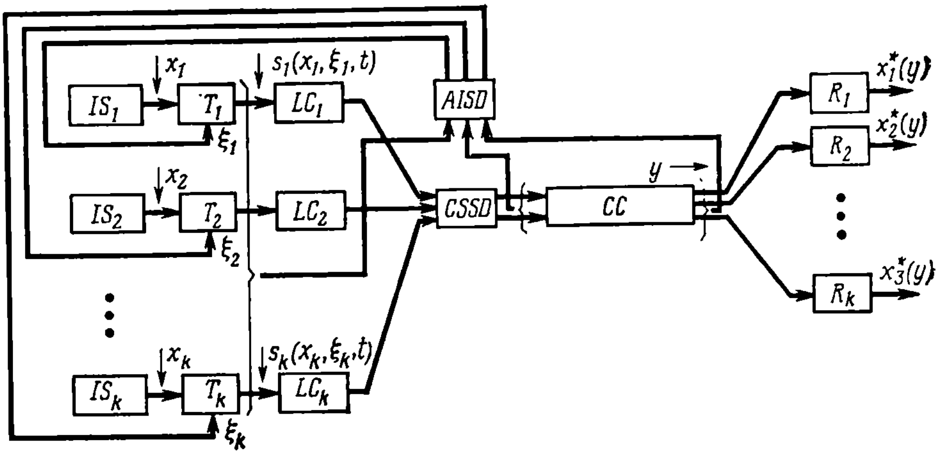

Fig. 8.1. Block diagram of remote multiaccess system: IS—information source; T—transmitter; LC—local channel; CC—common channel; R—receiver; CSSD—channel signal shaping device; AISD—auxiliary information shaping device

and so forth. In practice, one of these levels of the hierarchy is taken to be a base one and its constituent blocks are regarded as inseparable. In the process of transmission, such a block acquires auxiliary information (e.g., addresses) with the result that an information unit, termed *packet*, is shaped. As a rule, the "pulse ratio" of the packet flow is very small, of the order of 10^{-3} and less. However, primary communication channels are used most efficiently if signals being transmitted over them arrive uniformly with a pulse ratio close to unity. Such a misfit between the information flow characteristics and desired properties of signals in channels might, to a certain degree, be eliminated with a buffer memory. Suppose that in transmitters (network nodes) there are buffers where packets may be stored. At the same time the matching of behaviour of information sources with that of channels may be partially simplified through the use of one channel to communicate information from a multitude of sources. This possibility is one of distinctions of a computer-communications network and remote multiaccess systems.

8.2. Remote Multiaccess Systems

Introductory remarks. Fig. 8.1 presents a block diagram of the main system enabling a common channel to be used for several sources. Signals from individual sources, after being transmitted over a local channel, are introduced into a signal shaper at the common channel input, that transforms the incoming signals and selects a channel common signal. A typical example of such a system is a satellite

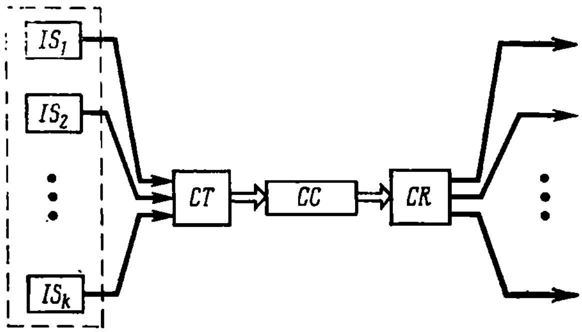

Fig. 8.2. Multiaccess system with central transmitter: CT—common transmitter; CC—common channel; CR—common receiver

repeater system. Another example is a system where signals from many sources arrive over local cable lines at the input of a long common coaxial or radio relay channel.

It is apparent that the common channel may be best utilized if information from individual sources is formed in a common buffer to be transmitted as a certain unit with the use of known methods of modulation, coding, and sometimes of feedback, which allow the common channel to be efficiently used. This system, known as the system with a central transmitter, is presented in Fig. 8.2. In the general case, the application of this system involves great costs, and may be even impossible. Therefore, it would be advisable to transfer a part of operations associated with the generation of a common signal from the central transmitter to individual ones, leaving for the common channel input relatively simple operations only.

In the simplest case this may be a superposition of electromagnetic waves on the air, the efficiency of the system being the higher, the more auxiliary information on the system condition will be available to individual transmitters.

Classification of individual transmitter performance. When transmitters function independent of the condition of the whole system, the algorithm of their performance is said to be the algorithm of rigid operation. Most frequently it is realized through multiplexing with separation of the common channel into subchannels and assignment of a given subchannel to a definite source. But the separation of the common channel into subchannels is not indispensable in realizing the rigid operation algorithm. An illustration is the address-pulsed asynchronous systems with nonorthogonal signals.

In order that individual transmitter may be more efficient in the use of the common channel, it should be provided with auxiliary information on the state of the whole system, and more specifically, on the state of the common channel and other individual transmitters. Owing to the auxiliary information, the transmitter is capable of

adapting its transmission conditions to the state of the system. The signal communicated by the kth transmitter ($k = 1, 2, \ldots$) could then be written as $s_k(x, \xi, t)$ (see Fig. 8.1). Here the access algorithm $s_k(\cdot)$ is termed the alternative access algorithm. Using this algorithm, the common channel may be utilized both with separation and without separation into subchannels. With separation, the serial number of a subchannel assigned to a given transmitter may vary with the system state.

We consider in more detail the algorithm of alternative access without subchannel separation. More often than not only some of signal parameters vary. Two approaches to forming these parameters are possible, distributed and centralized. In the first case, each individual transmitter possesses auxiliary information and forms on its own the parameters of transmitted signal conditioned by the former. An illustration of such a system is a feedback system. Depending on the auxiliary information at the feedback channel output on the quality of signal at the common channel output, the transmitter either repeats the transmission of a given packet or begins to communicate the following one. Another example of a system without subchannel separation may be a system in which the transmitter, having received a packet from a source, estimates the state of the common channel, and having established that it is vacant, begins to transmit the packet to the channel.

With the centralized forming of signal characteristics, auxiliary information is forwarded to the control center which then sets these characteristics for all the individual transmitters. An example of this system is a system that interrogates in succession individual transmitters, whether or not they have packets for transmission. If the answer is positive, the transmitter is permitted to connect to the common channel. A more sophisticated form is the system with orders. Here the transmitters that have received packets from sources notify the control center to the effect. The control center in turn communicates to individual transmitters the transmission mode specified for them. This system is similar to the system with a central transmitter in Fig. 8.2.

Remote access algorithm. The algorithm of operation of individual transmitters is characterized by two groups of parameters, viz. by the parameters related to the transmission quality and costs. In the first group, the two major parameters are the degree of distortions and the time taken to communicate information within the system. As the first parameter, the probability P_{er} of erroneous reproduction of information in optimal reception is often selected, and for the second, the average time τ elapsed from the generation of information by the source to the instant of its arrival at destination. Typical second-group parameters are costs of the system as related to one packet of information transmitted and operational costs.

The exact determination of these parameters is a difficult problem. In the general case, the main parameter is chosen so that it characterize a certain transmission mode under certain conditions. Then an appropriate model is constructed that incorporates nondeterministic factors defining the behaviour of the whole of the system, and lastly, the dependence of the parameter selected on random factors is eliminated, but its dependence on common channel access procedure is retained.

We take by way of illustration the parameter characterizing the access algorithm of the kth transmitter in terms of transmission fidelity. When information x^* is communicated, and at the destination information x arrives, the quality of transmission is described by the loss function $R(x, x^*)$. We denote by $x^*(\cdot)$ the rule of information derivation by which the reproduced information $x^*(y)$ corresponds to the signal y at the channel output.

The loss $R[x, x^*(y)]$ depends on the initially unknown x, y and the reception rule $x^*(y)$. The parameter $\bar{\varepsilon}[s_k(\cdot)]$ pertaining to the access rule $s_k(\cdot)$ may be represented as

$$\bar{\varepsilon}[s_k(\cdot)] = \underset{x,\, y,\, x^*(\cdot)}{DEO} R[x, x^*(y)] \tag{8.1}$$

where DEO is the dependence elimination operation for the dependence on x, y, $x^*(\cdot)$. The selection of the form of this operation is governed both by the adopted model of nondeterminated factors and by the access procedure. If a model with full a priory information (Bayes) is used, then more often than not

$$\underset{x,\, y,\, x^*(\cdot)}{DEO} = \underset{x^*(\cdot)\, X\cdot Y}{DEO\cdot M} \quad \text{for} \quad x \in X \quad \text{and} \quad y \in Y \tag{8.2}$$

where M is the operation of statistical averaging. Another most common operation is

$$\underset{x,y,x^*(\cdot)}{DEO} = \underset{x^*(\cdot)}{DEO} \max_{x,y} \tag{8.3}$$

The selection of DEO is dependent on the possibility to affect the access procedure. If $x^*(\cdot)$ is selected unlimited, then

$$\underset{x^*(\cdot)}{DEO} = \min_{x^*(\cdot)} \tag{8.4}$$

If operations according to Eqs. (8.2) and (8.4) are used, then the parameter for the access procedure $s_k(\cdot)$ in terms of distortions is taken to be

$$\bar{\varepsilon}[s_k(\cdot)] = \min_{x^*(\cdot);X\times Y} MR[\tilde{X}, x^*(\tilde{Y})] \tag{8.5}$$

where $\tilde{X}$ and $\tilde{Y}$ are random processes yielding information on the signal at the channel output. If

$$R(x, x^*) = \begin{cases} 1, & \text{for} \quad x \neq x^* \\ 0, & \text{for} \quad x = x^* \end{cases} \tag{8.6}$$

i.e. it constitutes a simple loss function, then the parameter of Eq. (8.5) is essentially the error probability P_{er} when using the optimal reception procedure.

We dealt as yet with a single procedure $s_k(\cdot)$. Denote the parameter for the set of these procedures

$$s(\cdot) = \{s_k(\cdot)\}, \; k = 1, 2, \ldots K \tag{8.7}$$

by $\bar{\bar{\varepsilon}}[s(\cdot)]$. This parameter is given by

$$\bar{\bar{\varepsilon}}[s(\cdot)] = \underset{k}{DEO}\, \bar{\varepsilon}[s_k(\cdot)] \tag{8.8}$$

A typical example of DEO is weighted averaging

$$\underset{k}{DEO}\, \bar{\varepsilon}[s_k(\cdot)] = \sum_{k=1}^{K} \beta_k \bar{\varepsilon}(s_k(\cdot)] \tag{8.9}$$

where β_k is the weight coefficient.

In communications between computers the value of parameter pertaining to distortions, e.g. error probability, is normally specified. Therefore we consider a second parameter of transmission quality, i.e. the parameter $\bar{\tau}[s_k(\cdot)]$, characterizing the time of packet passage through the system. We assume that all the information sources, individual transmitters and local channels have similar properties. Then $\bar{\tau}[s_k(\cdot)] = \bar{\tau}[s(\cdot)]$, therefore we denote this parameter τ for brevity.

We turn now to parameters characterizing the transmission costs. The major contribution is from the construction of the common channel. The parameter describing the costs of transmission of a unit information is

$$\nu = R_{com}/C_{com} \tag{8.10}$$

where R_{com} is the transmission rate over the common channel of the total information flow from all the sources; C_{com} is the common channel capacity.

We will refer to the parameter ν as the channel utilization efficiency (see Sec. 3.4). So, for a binary disturbance-free channel functioning with interruptions, the parameter ν is here as a pulse ratio for the total signal flow.

Properties of some remote multiaccess algorithms. We examine the function $\tau(\nu)$, in particular its being influenced by auxiliary infor-

mation on the state of the system, ξ, arriving at individual transmitters.

Let us take first the algorithm of rigid operation with subchannel separation. The system with such an algorithm is equivalent to a certain totality of independent information transmission systems. Denote by

$$\nu_k = R_k/C_k \tag{8.11}$$

the efficiency of utilization of the kth subchannel, where C_k is its capacity; R_k is the information transmission rate over that subchannel. Denote by

$$\zeta_c = \sum_{k=1}^{K} C_k/C_{com} \tag{8.12}$$

the efficiency of utilization of the common channel capacity. The difference $C_{com} - \sum_{k=1}^{k} C_k$ is here a part of capacity taken for the purpose of organization of subchannel operation and, in particular, for the separation of protective time or frequency zones. By Eq. (8.9)

$$\nu_k = \nu_1 = \text{constant} \tag{8.13}$$

We obtain from Eqs. (8.8), (8.11), and (8.12)

$$\nu = \nu_1 \zeta_s \tag{8.14}$$

The only way of increasing ν_1 is a reduction in subchannel capacity, which is equivalent to a longer accumulation time of the elementary signal. Clearly, this requires a buffer storage in each individual transmitter. As is known [473], a typical example of the variation of $\bar{\tau}$ with ν_1 is

$$\bar{\tau} = \bar{U}/(1 - \nu_1) \tag{8.15}$$

where $\bar{U}$ is the average packet length. As $\zeta_s < 1$, the variation of $\bar{\tau}$ with ν has the behaviour as shown in Fig. 8.3, where

$$\nu_0 = \zeta_s \tag{8.16}$$

We now consider the rigid operation algorithm without subchannel separation. An illustration of this system is the system with pseudorandom, continuous or pulsed, signals. The major distinction of this system is that signals from various transmitters in the common channel give rise to mutual interference. The system is described by the parameter — the efficiency of utilization of common channel capacity — defined by Eq. (8.12), where C_1 signifies the capacity of the channel all the way from an individual source to the information destination. Examples of the capacity utilization efficien-

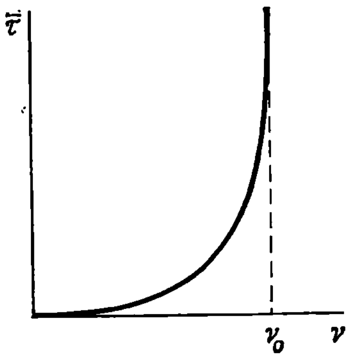

Fig. 8.3. Average time of packet passage through the system

cy are provided in [473], where it has been proved that the efficiency is dependent on the number of individual transmitters K and the number of degrees of freedom of the signal $N = 2TB$, where T is its duration; B is the bandwidth occupied by the signal. This dependence is illustrated in Fig. 8.4*a*, *b*, where ρ denotes the signal-to-noise ratio. It follows from the theory of error control coding that the error probability in optimal detection of sufficiently long signals is given by

$$P_e = A \exp[-T\alpha(C_1, R_1)] \tag{8.17}$$

where T is the duration of signals; A is a coefficient, R_1 is the rate of information transmission from an individual source; $\alpha(C_1, R_1)$ is a coefficient dependent on signal behaviour and noise level in the channel. It has been shown for typical cases that $\alpha(C_1, R_1) \to 0$,

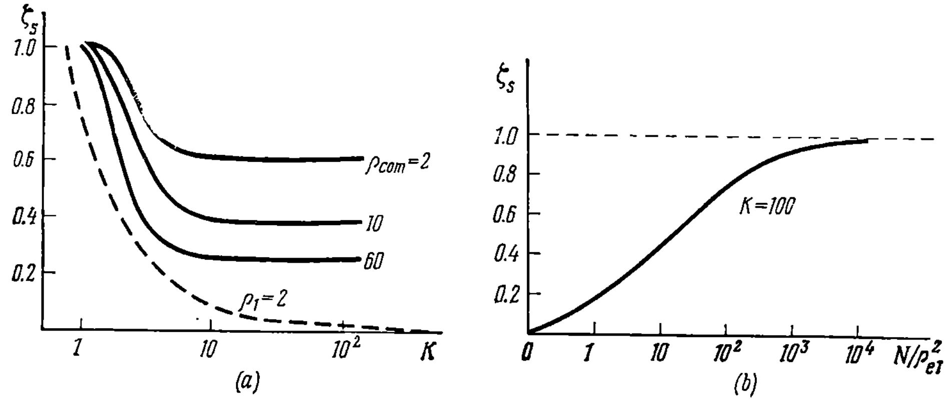

Fig. 8.4. Efficiency of common channel capacity utilization: ρ_{com}—power ratio of total group signal to noise in common channel; ρ_{e1}—ratio of signal energy to noise spectral density

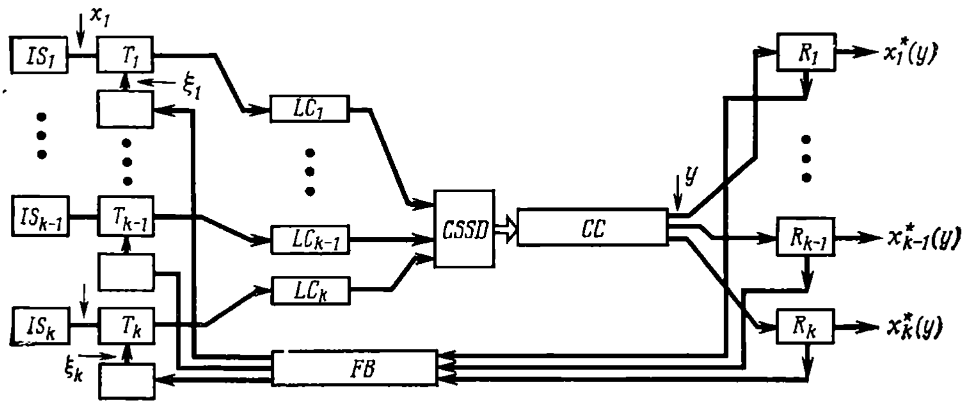

Fig. 8.5. Block diagram of remote multiaccess system with feedback (for symbols see Fig. 8.1)

if $R_1 \to C_1$. It follows from the above argument that to obtain a predetermined probability $P_{er} \ll 1$, it is necessary that $T \to \infty$ for $R_1 \to C_1$. This is equivalent to

$$\nu \to \zeta_s \tag{8.18}$$

As the signal is transmitted and received as a unit, the average time from the instant the information is received from the source to the instant the signal carrying it is sent to the channel is $T/2$, and the reception of the signal takes the time T, then the time of passage of information through the system will be

$$\tau \approx 3T/2 \tag{8.19}$$

Hence the variation of τ with ν has the form as indicated in Fig. 8.3 ($\nu_0 = \zeta_s$). It is readily seen that a similar dependence may be obtained using not block but recurrent codes.

Consider now a system with feedback. Over the feedback channel in the system of Fig. 8.5, auxiliary information ξ is communicated about the quality of the signal received. In the simplest case ξ is binary information that takes one of the two forms: ξ_A—the received signal is satisfactory, ξ_D—the received signal is rejected. If the auxiliary information ξ_D is received, the packet is retransmitted. As a result, two flows of packets pass over the common channel: packets transmitted satisfactorily, and those causing the information ξ_D to be sent. We denote by ν^* the pulse ratio of the packet flow with satisfactory quality, and by ν_{com} the pulse ratio of the flow for all the packets in the common channel. A typical variation of ν^* with ν_{com} is provided in Fig. 8.6 (the derivation of this and similar variations is to be found in [308, 474]). It can be shown that if $\nu_{com} > \nu_{max}$ (where ν_{max} is the value of the parameter ν_{com} for

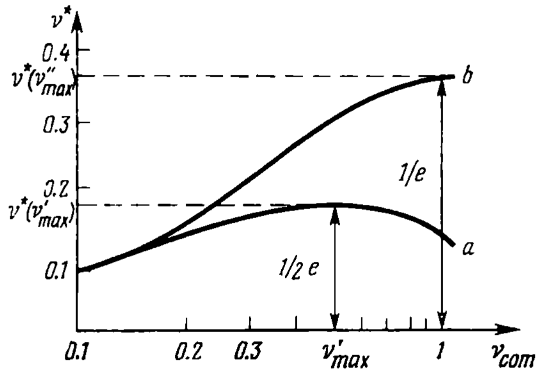

Fig. 8.6. Pulse ratio for packet flow

which ν^* attains a maximum), then in the system the number of repeated packet transmissions (requests for retransmission) grows in an avalanche manner. With earlier nomenclature, $\nu^* = \nu$. Thus, the average time ν from the instant of appearance of a packet to the instant of its successful reception is here also characterized by the dependence of Fig. 8.3, where $\nu_0 = \nu^* (\nu_{max})$.

The efficiency of the system might be improved by introducing markers indicating equal time intervals that permit packet input within the specified interval only. Fig. 8.6 shows a typical variation of ν^* with ν_{com} for such a system (curve *b*). The character of the behaviour of $\bar{\tau}$ (ν^*) remains much as in Fig. 8.3 (see also [410, 424, 440, 474]).

Now we consider a polling system. When the system starts operating, the control center sends the interrogating signal ξ'_1 to the transmitter with serial number m_1. The transmitter sends auxiliary binary information ξ''_A if it has working information (one or more packets designed for transmission), or ξ''_D if it has no such information. The information ξ''_A is followed by working information, and thereafter auxiliary information indicating that the transmission of the working information from the transmitter m_1 is over. Having received auxiliary information ξ''_D or ξ''_A, the control center sends a signal ξ'_2 to the transmitter m_2, and so on. A detailed treatment of this system is to be found in [384, 419]. The dependence of $\bar{\tau}$ on ν_{com} is as in Fig. 8.3, but instead of ν_0 a parameter should be used which depends on the structure of signals bearing auxiliary information. The just-considered system may be called the centralized polling system. A decentralized form is also possible, in which the auxiliary information ξ''' and ξ''_D is transmitted not to the control center but to the next individual transmitter. The above system is fairly similar to the system with testing for the condition of the common channel. In the latter the transmitter, havin established that another transmitter

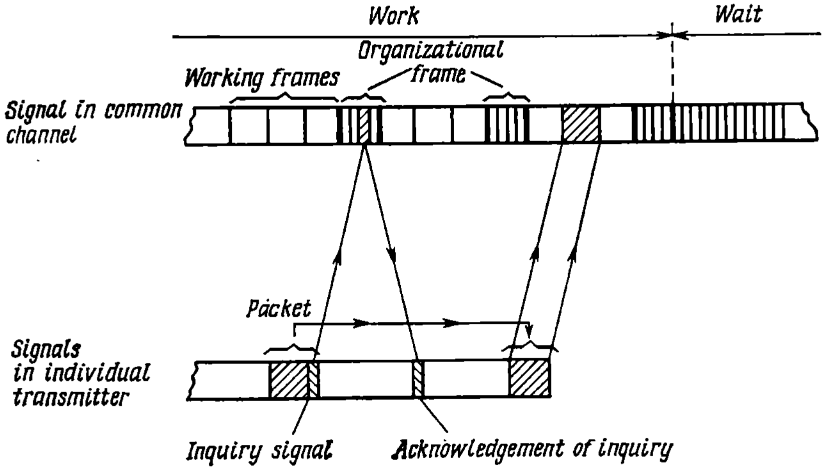

Fig. 8.7. Signal structure in call system

has finished its session, starts transmitting its information to the common channel. A disadvantage of this system, as compared with the system with interrogation, is the presence of a time interval during which two or more transmitters may unawares begin their session simultaneously. This may be obviated using feedback, but this time through longer $\bar{\tau}$. The discussion of these systems is available in [411, 474].

Lastly, we consider the call system in which a subchannel is allocated from the common channel for individual transmitters to send auxiliary information about the packet number being on the waiting list for transmission. Using this auxiliary information, the control center draws up a time table transmitted over return channels to all the individual transmitters. A typical structure of signals in this system is presented in Fig. 8.7.

Clearly, the call system behaves in much the same manner as the system of Fig. 8.2 with a common buffer in the central transmitter. In fact, the call system has one common waiting list but the waiting packets are in various places. The delay time for a system with a waiting list may be determined according to Eq. (8.15) putting $\nu_1 = \nu$, for $\nu < \nu_{call}$ and

$$\nu_{call} = 1 - C_{call}/C_{com} \tag{8.20}$$

where C_{call} is a part of channel capacity occupied by call signals. A detailed analysis of call systems is to be found in [462, 474].

The delay in the systems considered is due to fundamentally different causes. So, in the rigid access system without subchannel separation, the delay is caused by the required waiting for transmis-

sion and reception of a long code sequence, and in the feedback system, by the required waiting for successful transmission. In the system with separated subchannels, packets are on independent waiting lists; in the interrogating system, on interrelated waiting lists, and in the call system, on one common waiting list. Nonetheless, in all the discussed cases the variation of $\bar{\tau}(\nu)$ shows similar behaviour. In addition, increased auxiliary information not only results in shorter $\bar{\tau}$, but in longer ν_0 as well.

The behaviour of $\bar{\tau}(\nu)$ discussed above for various forms of auxiliary information may be regarded as similar to the variation of reception error probability with signal-to-noise ratio in the channel for various degrees of functional coordination of the receiver with the transmitter. To the system with central transmitter or call system there corresponds the system with completely coherent operation of the transmitter and receiver.

Thus, the information transmission time for the system may be shortened relatively easily at the expense of channel efficiency. Ople timal selection of the system type and its parameters is only possible through a painstaking analysis of the system with due regard for the common channel costs, costs of the subsystem of shaping and transmitting the auxiliary information, and also for the required quality of transmission of working information.

8.3. Networks

Introductory remarks. The construction of communication networks for distributed computer systems is one of the most important present-day practical tasks. Economic aspects of these networks are discussed in [307]. A new element here, as compared with the above systems with one common channel, is the possibility of route selection to transmit signals in the network. Clearly, the information transmission over the network may use feedback, polling, and calls. Although these methods have found practical application in certain real networks, but the theory of their use in involved systems is still at its infancy. We suppose that the nodes in the network may equally well be transit nodes, information sources, and also information sinks. In order to set off these two latter functions, we will refer to the node as the pole (source pole, sink pole).

Classification of routing techniques. By the route selection technique (routing technique) we will call the technique by which information from a source pole will be directed to a sink pole. These techniques may be classified based on various factors. We begin with the main factor connected with the question of whether information in passing through the network should be regarded as the indivisible whole. Granted that it be the case, the information upon arrival at

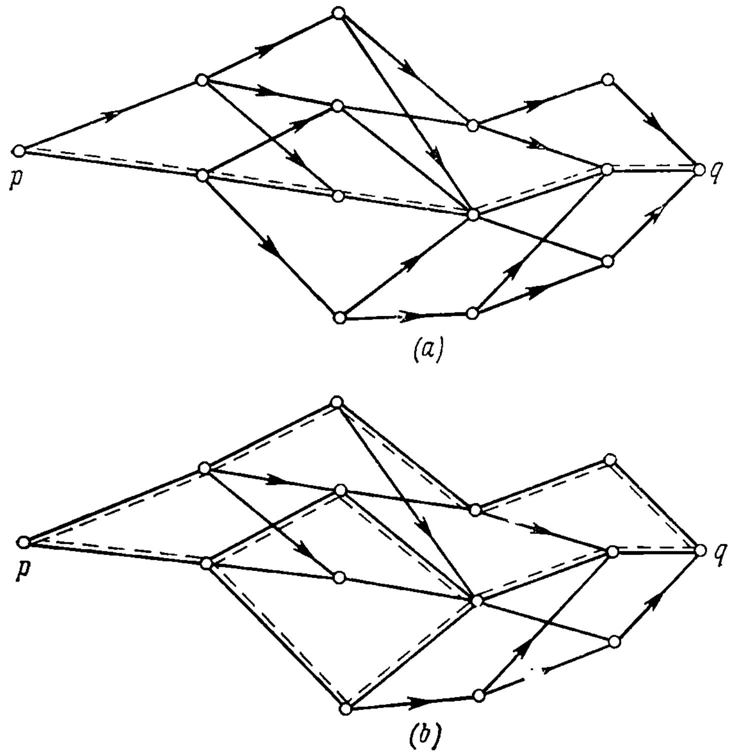

Fig. 8.8. Routes: (*a*) without branching, (*b*) with branching

a given information node is fed as a unit to one of channels originating at this node. Thus the route to be followed by information through the network has no branching (Fig. 8.8). Such a routing technique is termed the selection of route without branching. If further division of information is possible, then it might appear that it pays to send such smaller parts over different channels originating from the node. In that case the route taken by the information through the network may have branchings, and the appropriate routing technique is called the selection of route with branching.

Information may be broken down into smaller parts in various ways. Most frequently the block is subdivided into subblocks. To each of subblocks data should be added which allow the block to be reassembled from subblocks again after their arrival at the sink pole.

A second property of routing technique is associated with the method for building a route of channels connecting the nodes. This may be achieved in two ways, rigid and variable. In the first technique the route is not subject to any change. In that case the technique is known as the selection of a fixed route. If the route may change

depending on the network state, then the rule of its selection is said to be the selection of a variable route. As theie is always a tendency for the route to be adapted to the network state, the selection of a variable route may also be called the adaptive routing. The pairs of route properties—with branching and without branching, fixed and variable—are independent. Therefore any combinations of them are possible.

As has been indicated in Sec. 8.2, information, and in particular the information communicated between computers, shows the block-hierarchical structure. The routing rule for blocks at a given hierarchical level determines the routing technique for sequences of these blocks, and also the route selection for higher-level blocks. The routing technique for blocks and related routing technique for sequences of blocks may belong to different classes. Thus, if for blocks the variable routing without branching is used, then for the sequence of blocks the routing with branching may be selected.

The selection of variable route is of especial practical value. The description of these techniques is to be found in [369, 465. 474, 483]. In case of variable routes, a remarkable part is played by the behaviour of auxiliary information about the network state. Using the control theory therminology, this auxiliary information may be called the data for system identification. In a real network, this auxiliary information is not always completely actual not only because it arrives from space-distributed nodes, but because the node polling takes place at certain instants of time only. The difference in time between the instant at which auxiliary information can be really used in decision-making and the instant which it characterizes is said to be the degree of nonactuality. Another parameter to characterize auxiliary information is the distance between the node characterized and the node where a decision is taken as to the selection of route. We arbitrarily call this distance the "range" of auxiliary information. The channel capacity drops if a large nonactuality of auxiliary information is allowed and it grows with the range of auxiliary information. As a part of capacity alloted to transmit auxiliary information is limited, the degree of its nonactuality and its range are interrelated as is shown in Fig. 8.9.

Consider two cases. In one the nonactuality is large as compared with the mean time of packet transmission through the network. In that case it is advantageous to define all of the route from the source pole to the sink pole. We will call such a procedure the pole-to-pole routing. Further, in the case considered it would be good practice to have a center located at one of the nodes to which all the auxiliary information is communicated and at which all the decisions are taken about the routing, to be transmitted then to all the nodes. Such a system is said to be the system with centralized routing decison-making.

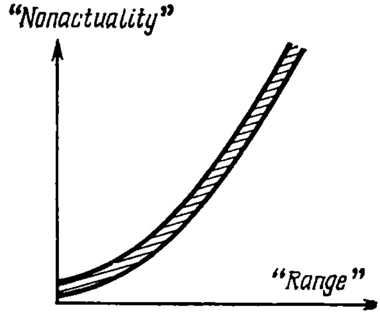

Fig. 8.9. Typical variation of "nonactuality" with "range" of auxiliary information

In the second extreme case the nonactuality is small as compared with the mean time of packet transmission through the network. Here it would pay not to take decisions about all the route from pole to pole, but after the packet has achieved a given node, to take decision as to what adjacent node the packet should be directed. We will call such a procedure the node-to-node routing. As is seen in Fig. 8.9, for small nonactuality, the range of auxiliary information is not long. Therefore it is advisable to communicate the auxiliary information to all the nodes and take decisions about routing in situ at a given node. We will refer to this procedure as distributed routing decision-making.

Thus, two sets of described properties are most expedient. The first one: large nonactuality, long "range", pole-to-pole routing, centralized routing decision-making; the second one: small nonactuality, short "range", node-to-node routing, distributed routing decision-making. Clearly, other forms are also possible.

Between the "pole-pole" and "node-node" procedures, intermediate routings are also possible. The simplest of them consists in dividing the network into regional subnetworks and using within the subnetworks the "node-node" procedure, and between the subnetworks, the "pole-pole" procedure. Such an approach is discussed in [405, 412].

The centralized and distributed routing decision-making may be combined not only in "space". For example, when the long range auxiliary information is considered reliable, the centralized "pole-pole" procedure may be utilized. Should this information be estimated as unreliable, then it is wise to employ a distributed "node-node" procedure based on some more reliable shorter range auxiliary information. Such a form of the procedure, termed Δ-routing, is considered in [465].

An obvious deficiency of the distributed routing is its being based on short range auxiliary information. One of the corollaries of this

method is the fact that the network load grows, and the transmission quality deteriorates faster than with the centralized routing. This may be remedied by using techniques of preventing local overloads, a review of which is available in [464]. The following procedure is a typical one:

1. A packet is allowed entering the network, if at the node into which it is introduced from the source there are "entrance tickets" (ET), if only one.
2. The number of all ETs in the network is fixed.
3. The entrance ticket is introduced into the packet heading.
4. The entrance ticket is freed after the packet has arrived at a sink pole.
5. The freed ET stays at this node or somehow is transferred to the network.

Clearly, the principle of the method is in the introducing of additional auxiliary information on the state of all the network.

We now consider the selection of variable route "node-node" in more detail. If the selection of the channel into which a packet is fed is only dependent on (a) sink pole, (b) source pole, (c) auxiliary information on the network state, then the routing technique is said to be deterministic. If, however, this selection is only dependent on (b) and (c), but is independent of (a) and the route already covered, the procedure is referred to as the memoryless routing. The class of procedures with memory is rather wide. An example is the procedure eliminating any possibility of packet circulation over closed loops. With such a routing, the serial numbers of just-passed channels are stored in the packet heading. These numbers enable the packet that got into a closed loop to be identified and to change its route accordingly. As a last resort, the packet may be discarded.

Routing parameters. In order that different routing procedures may be compared and appropriate optimization problems formulated, parameters should be introduced to characterize these procedures. It is assumed that these parameters may be divided into two groups by the transmission quality and costs, defined in exactly the same way as the parameters of the remote access procedures. We introduce first a parameter characterizing the concrete communication of packet over a concrete route at a given state of the network. Next, using an appropriately selected operation of dependence elimination from nondeterminated factors we arrive at a parameter characterizing the routing procedure. We illustrate the procedure by a worked example.

Denote by $\mathcal{L}_{pq}(\cdot)$ a determinated procedure of route selection from a pole p to a pole q. If ξ denotes an accessible auxiliary information, then $\mathcal{L}_{pq}(\xi)$ is a route followed by the information packet. Further, we designate the set of these routing procedures as $L(\cdot) = \{\mathcal{L}_{pq}(\cdot)\}$, for $p, q = 1, 2, \ldots, N$, where N is the number of

network nodes. The parameter characterizing $L(\cdot)$ as regards the time of passage of the packet through the network is determined in three steps. In the first step a parameter $\tau[D(p, q)]$ is introduced to describe the transmission time of the packet over a given route $D(p, q)$ from the pole p to the pole q. We write

$$\tau[D(p, q)] = \sum_{(w, v)\in D(p, q)} \tau(w, v) \tag{8.21}$$

where $\tau(w, v)$ is the transmission time for the packet over the channel (w, v); and sum up it over all the channels constituting the route $D(p, q)$. We then denote by V the set of numbers defining the network state (number of packets stored in buffer memories; numbers characterizing the state of communication channels). To indicate the dependence of $\tau(w, v)$ on V we write $\tau(w, v, V)$. The time of transmission of the packet by the route according to the routing procedure $\mathscr{L}_{pq}(\cdot)$ for a given ξ will be

$$\tau[\mathscr{L}_{pq}(\xi), V] = \sum_{(w, v)\in\mathscr{L}_{pq}(\xi)} \tau(w, v, V) \tag{8.22}$$

In the second stage we eliminate the dependence on V and ξ to obtain a parameter defining the procedure as a whole. Using for *DEO* the operation of statistical averaging, we obtain

$$\bar{\tau}[\mathscr{L}_{pq}(\cdot)] = \underset{V\times\theta}{M} \sum_{(w, v)\cdot\mathscr{L}_{pq}(\theta)} \tau(w, v, V) \tag{8.23}$$

where the averaging is performed over all the states of the network and over all the volume of auxiliary information. The symbols V and θ designate random quantities of processes, whose realizations are V and ξ. In the third stage, using the parameter $\bar{\tau}[\mathscr{L}_{pq}(\cdot)]$ as a basis, we determine the parameter $\bar{\bar{\tau}}[L(\cdot)]$ characterizing the set $L(\cdot)$. To this end it is necessary to eliminate the dependence on p and q. As a rule for this purpose weight coefficients $\beta(p, q)$ are utilized, pertaining to the transmission rank from p to q. We thus have

$$\bar{\bar{\tau}}[L(\cdot)] = \sum_{p, q} \beta(p, q)\, \bar{\tau}[\mathscr{L}_{pq}(\cdot)] \tag{8.24}$$

If G_{pq} stands for the information flow intensity from p to q, then as a weight coefficient, the relative intensity is often used

$$\beta(p, q) = G_{pq}/\sum_{p, q} G_{pq} \tag{8.25}$$

For $\beta(p, q)$ thus determined, after some manipulation, we obtain

$$\bar{\bar{\tau}}[L(\cdot)] = \frac{1}{\sum\limits_{p, q} G_{pq}} \sum_{w, v} \beta(w, v)\, \Phi\, w, v)\, \bar{\tau}(w, v) \tag{8.26}$$

where $\Phi(w, v)$ is the average intensity of flow of all the packets in the channel (w, v); $\tau(w, v)$ is the average transition time of packets over this channel.

The parameter given by Eq. (8.24) does not consider the spread of parameters $\bar{\tau}[\mathscr{L}_{pq}(\cdot)]$. These spreads may be taken into account through the use of the parameter

$$\bar{\bar{\tau}}_{\gamma}[L(\cdot)] = \{\sum_{p, q} \beta(p, q)\bar{\tau}^{\gamma}[\mathscr{L}_{pq}(\cdot)]\}^{1/\gamma} \tag{8.27}$$

where $\gamma \geqslant 1$ is the auxiliary coefficient. A limiting case is of interest where $\gamma \to \infty$:

$$\bar{\bar{\tau}}_{\infty}[L(\cdot)] = \max_{p, q} \bar{\tau}[\mathscr{L}_{pq}(\cdot)] \tag{8.28}$$

A more detailed discussion of methods for determining the figures of merit for access rules and practical examples of derivation of these figures are to be found in [474].

Routing optimization. Other parameters describing the routing procedure may be obtained in much the same way as $\bar{\tau}$ or $\bar{\bar{\tau}}$. One of these parameters is normally taken to be a criterion, whereas others are fixed or specified within certain limits. Denote by $\bar{\varepsilon}[\mathscr{L}_{pq}(\cdot)]$ or $\bar{\bar{\varepsilon}}[L(\cdot)]$ the parameter selected to be the criterion. There exist two types of optimization problems. Problems of the first kind arise where only the procedure $\mathscr{L}_{pq}(\cdot)$ is considered; the second kind of problems occurs where all the set $L(\cdot)$ is considered. The problem of the first kind may be formulated as follows:

> *in a class of routing procedures defined by conditions C it is required to find a procedure $\mathscr{L}^0_{pq}(\cdot)$ for which the criterion $\bar{\varepsilon}[\mathscr{L}_{pq}(\cdot)]$ attains an extreme value (minimum or maximum depending on the sense of $\bar{\varepsilon}$).* (8.29)

We denote this problem by $OP\, \mathscr{L}_{pq}(\cdot), \bar{\varepsilon} \mid C$, where $\mathscr{L}_{pq}(\cdot)$ is a "variable"; $\bar{\varepsilon}$ is a criterion, and C are conditions. In a similar manner, proceeding from the criterion $\bar{\bar{\varepsilon}}$ we formulate optimization problems for $L(\cdot)$.

To begin with, consider the solutions to $OP\, \mathscr{L}_{pq}(\cdot)$ without conditions, assuming that:

(1) the criterion takes the form

$$\varepsilon(D) = \sum_{(w, v) \in D} \varepsilon_1(w, v) \tag{8.30}$$

(2) the parameter $\varepsilon_1(w, v)$ characterizing a single channel is independent of the condition of the network. The quantity $\varepsilon_1(w, v)$

may be thought of as the length of the channel (w, v). We designate

$$d(w, v) = \varepsilon_1(w, v) \tag{8.31}$$

then

$$d[D(p, q)] = \sum_{(w, v) \in D(p, q)} d(w, v) \tag{8.32}$$

has the sense of route length $D(p, q)$. We minimize $\bar{\varepsilon}[\mathcal{L}_{pq}(\cdot)]$ when the packets are directed along the route $D_0(p, q)$ that is the shortest as defined by Eq. (8.32). Then, under the both assumptions the solution to $OP\,\mathcal{L}_{pq}(\cdot)\,\varepsilon$ is the procedure:

a packet should be directed from the pole p to the pole q along the route, the shortest in the sense of Eq. (8.32) (8.33)

We suppose that: (1) the state of the network may be thought of as a multidimensional random quantity (process); (2) the criterion has the form of Eq. (8.30); (3) the parameter for a single channel is dependent, in a general case, on the state of the network V. To emphasize the last assumption we introduce $\varepsilon_1(w, v, V)$. Let us define

$$\varepsilon[D(p, q), V] = \sum_{(w, v) \in D(p, q)} \varepsilon_1(w, v, V) \tag{8.34}$$

$$\bar{\varepsilon}[\mathcal{L}_{pq}(\cdot)] = \underset{V \times \theta}{M}\, \varepsilon[\mathcal{L}_{pq}(\theta), V] \tag{8.35}$$

where the averaging is carried out over all the pairs of V, θ. We introduce further the conditional mean

$$\bar{\varepsilon}[D \mid \xi] = \underset{V \mid \xi}{M}\, \varepsilon(D, V) \tag{8.36}$$

We have

$$\bar{\varepsilon}[D \mid \xi] = \sum_{(w, v) \in D} \bar{\varepsilon}_1[(w, v) \mid \xi] \tag{8.37}$$

where

$$\bar{\varepsilon}_1[(w, v) \mid \xi] = \underset{V \mid \xi}{M}\, \varepsilon_1[(w, v), V] \tag{8.38}$$

Using the formula for the conditional averaging, we write Eq. (8.35) in the form

$$\bar{\varepsilon}[\mathcal{L}_{pq}(\cdot)] = \underset{\tilde{\theta}}{M}\, \bar{\varepsilon}[\mathcal{L}_{pq}(\theta) \mid \theta] \tag{8.39}$$

We define the channel length by the relationship

$$d(w,\ v) = \varepsilon_1 [(w,\ v) \mid \xi] \tag{8.40}$$

Then $\varepsilon (D \mid \xi)$ will be the smallest for the route $D_0 (p,\ q,\ \xi)$ that is the shortest in the sense of Eq. (8.40). Thus, the solution to $OP\, \mathcal{L}_{pq}(\cdot),\ \bar{\varepsilon}$ is the procedure

> *a packet should be directed from the pole p to the pole q along the route $D_0 (p,\ q,\ \varepsilon)$, the shortest in the sense of Eq. (8.40).* (8.41)

The use of this corollary presents substantial difficulties owing to the fact that in the general case the probabilities of network states are dependent on the routing procedure. Therefore to derive the procedure effectively, iterative techniques are required. And even with these methods, the computation of conditional probabilities of Eq. (8.36) is much of a problem. With this in mind, it is expedient to make use of the general adaptation principle. In the present case, this results in the following modification of the procedure of Eq. (8.33):

> *a packet is directed along the route $D^* (p,\ q)$, the shortest as far as the channel length is concerned, that is defined by the relationship $d (w,\ v) = \hat{\varepsilon}_1 (w,\ v,\ \xi)$* (8.42)

where $\hat{\varepsilon}_1 (w,\ v,\ \xi)$ *is an estimate* $\varepsilon_1 (w,\ v,\ \xi)$ *based on* ξ.

Clearly, the quality of this procedure is to a large measure conditioned by the estimate $\hat{\varepsilon}_1 (w,\ v,\ \xi)$ selected. It might be expected, drawing on the major principles of control theory pertaining to adaptive systems, that the adaptive procedure of Eq. (8.42) will not always be absolutely optimal. On the other hand, an adaptive procedure similar to Eq. (8.42) has been applied in practice in the well known network ARPA. The details of operation of this network, the treatment and measurements of its characteristics are to be found in [463, 474]. It has been shown for this network that the procedure of Eq. (8.42) is actually the best among many other routing procedures.

Behavior of optimal routing procedures. In addition to the structure of an optimal procedure, a knowledge of its quality is also important. With this aim in view, simulation techniques are finding ever increasing use (see, e.g. [405]), which, however, yield special results only. The quality of optimized procedures could only be calculated analytically with simplifying artificially chosen assumptions. One example is the variation of average time of packet transmission through the network with Δ that has been derived in [475] and presented in Fig. 8.10. As Δ characterizes the "range" of auxiliary information, the result may be interpreted in the sense that where the same channels in a network are utilized to transmit working and auxiliary information, there exists an optimal "range" Δ of auxiliary information.

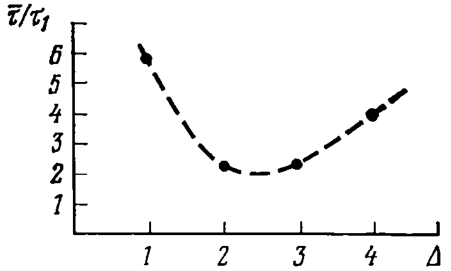

Fig. 8.10. Variation of average time of passage through network with Δ

Consider in more detail the method whereby optimal routing procedures may be synthesized in general and their behaviour evaluated. This method, in essence, consists in dealing not with individual packets, but with averaged flows. We shall dwell first on the second type of averaging. Averaging is possible both over short and long time intervals. The assumption that the averaging is carried out over long intervals is equivalent to assuming that the flow is stationary with the result that its parameters are independent of time. Such a flow is described in terms of the intensities G_{pq} of flows coming from outside to the pole p, which are to be directed to the pole q, and also the intensity $\Phi_{pq}(w, v)$ of the flow component sent from p to q along the channel (w, v). We then denote $\Phi = \{\Phi_{pq}(w, v)\}$, where p, q, w, $v = 1, 2, \ldots, N$. Knowing the probabilistic behaviour of packet flows fed to the network and the routing procedure, all the components of the set Φ can be determined. There is no unique way back; given the set Φ, in the general case, it is impossible to determine the routing procedure for the packets in the network. But in an important special case, where the flow from p to q shows on branching, the route followed by the flow from the pole p to the pole q defines uniquely the route for an individual packet. Should the averaged flow have a branching, the routing procedure is obtained heuristically.

In order to evaluate the routing procedure relying on the behaviour of the average flow, a quality criterion for the average flow related to the figure of merit of the routing procedure shall be introduced. This criterion may be defined, for instance, in terms of the average time of packet travel through the network. We will refer in what follows to such a quality criterion of average flow as $\bar{\varepsilon}'(\Phi)$. It follows from the correlation of these criteria that the routing procedure, being optimal in the sense of the criterion $\bar{\varepsilon}$, results in a flow optimal in the sense of the criterion $\bar{\varepsilon}'$.

It should be borne in mind in formulating the optimization problem for the set Φ that its components are independent. These should be subject to three constraints:

a constraint caused by the continuous nature of the flow at each node w,

$$\sum_{(w,\, v)\in K_{out}} \Phi_{pq}(w, v) - \sum_{(u,\, v)\in K_{in}} \Phi_{pq}(u, v) = \begin{cases} G_{pq}, & v=p \\ 0, & v\neq p, \quad v\neq q \\ -G_{pq}, & v=q \end{cases} \tag{8.43}$$

where K_{in} is the set of channels leading into the node w; K_{out} is the set of channels leading out of the node w. We will denote this constraint $A_1(G)$, where $G = \{G_{pq}\}$, $p, q = 1, 2, \ldots, N$ is the set of external intensities; a constraint due to capacity of channels

$$\sum_{p,\, q} \Phi_{pq}(w, v) \leqslant C(w, v) \tag{8.44}$$

We will designate this constraint by $A_2(C)$, where $C = \{C(w, v)\}$ with $w, v = 1, 2, \ldots, N$ is the set of capacities of channels;

a condition of nonnegativeness

$$\Phi_{pq}(w, v) \geqslant 0 \tag{8.45}$$

which we will denote by A_3.

The problem of optimization of the set Φ may be formulated as follows: it is required to determine a set Φ such that minimizes the criterion $\varepsilon'(\Phi)$ subject to the constraints $A_1(G)$, $A_2(C)$, A_3. We name this problem $OP\,\Phi$, $\varepsilon' \mid A_1(G)$, $A_2(C)$, A_3. The optimal flow that is the solution to this problem is dependent on the set of intensities of external flows G. In actual practice, to arrive at an optimal routing procedure based on the optimal flow, the time axis is divided into equal intervals. In the (n-1)th interval we estimate the average external intensities G_{pq}; and in the nth interval we apply the optimal routing procedure corresponding to the optimal flow derived from these estimates, and estimate external intensities to be used in the nth interval, and so forth.

Not infrequently $\varepsilon'(\Phi) \to \infty$, when Φ tends to the limit of the area subject to the constraint $A_2(C)$. Then ε plays the role of the loss function and $OP\,\Phi$, $\varepsilon' \mid A_1(G)$, $A_2(C)$, A_3 is reduced to $OP\,\Phi$, $\varepsilon' \mid A_1(G)$, A_3.

We will suppose that, as earlier [see Eq. (8.30)], the figure of merit of the flow looks like

$$\varepsilon'(\Phi) = \sum_{(w,\, v)} \varepsilon_1[\Phi(w, v)] \tag{8.46}$$

where

$$\Phi(w, v) = \sum_{p,\, q} \Phi_{pq}(w, v) \tag{8.47}$$

We have for the route D

$$d(D) = \sum_{(w,\, v) \in D} [d\varepsilon_1(w, v)/d\Phi(w, v)] \tag{8.48}$$

The shortest route here $D_0^{(l)}(p, q)$, $l = 1, 2, \ldots, H_{pq}$ is such that

$$d[D_0^{(l)}(p, q)] = \Lambda_{pq} \tag{8.49}$$

where Λ_{pq} are constants, and for any route D

$$d[D(p, q)] \geqslant d[D_0^{(l)}(p, q)] \tag{8.50}$$

It is shown easily (see, e.q. [474]) that the solution to $OP\,\Phi$, $\varepsilon' \mid A_1(G), A_3$ is a set such that nonzero components of the flow follow the shortest routes only.

There exists a class of algorithms based on the notion of the flow deviation [368, 474]. Algorithms are also convenient which use the gradient projection [471, 474].

A common property of the above algorithms is that all their components $\Phi_{pq}(w, v)$ are calculated simultaneously and approached in a similar way. When used in real time, they are suitable to work out the route "pole-pole". In order to derive algorithms suitable for "node-node" routing procedures, flow components are to be isolated which originate from a given node. This may be accomplished by substituting for the intensities $\Phi_{pq}(w, v)$ of the flow, going from the pole p to the pole q and coming out of the node w into the channel (w, v) its portions divided by the intensity of the whole flow originating at w. We then suppose that the routing procedure is a memoryless procedure. In this case, instead of $\varphi_{pq}(w, v)$ the symbol $\varphi_q(w, v)$ can be utilized. Further, we introduce the quantity

$$d_q(w, v) = d\varepsilon/d\varphi_q(w, v) \tag{8.51}$$

that may be regarded as the channel length originating at the node w. To such channels $(w, v_0^{(l)})$, $l = 1, 2, \ldots, H_{wq}$, for which $d_q(w, v_0^{(l)}) = \Lambda_{ql}$ [whereas for other channels $d_q(w, v) \geqslant d_q(w, v_0^{(l)})$], we will refer as the shortest. We next denote $\varphi = \{\varphi_q(w, v)\}$, $q, w, v = 1, 2, \ldots, N$. The problem $OP_\varphi \mid A_{1\varphi}(G), A_{3\varphi}$ has been treated in [372]. It has been proved that its solution is a set for which nonzero flow components follow the shortest channels. This result suggests that routing procedures have good properties if they yield routes the shortest in the sense of estimates of the derivative of Eq. 8.51. The methods of obtaining these estimates are proposed in [311, 372, 472]. As indicated above, in the method using the averaging over large time periods, the time axis is broken down into intervals. In one interval, we estimate the average intensities of external flows, and in the following we use the routing procedure proceeding from the estimates obtained. In so doing, the auxiliary information about the state of the system is rather limited, so it is

assumed that in the next interval the intensities of external flows do not change (time-independent). The adopted model for the system will be more real if average flow intensities are introduced which may vary in time. Abiding by the deterministic model, we will assume that these time-dependent intensities are known in the interval (t_{n-1}, t_n). In actual practice, it corresponds to the case where for the beginning time interval the behaviour of intensities is predicted from the time history.

Assuming that the intensities vary with time, the state of the system can be given by $V(\cdot) = \{V_q(w, t)\}$, $q, w = 1, 2, \ldots, N$, and $t_{n-1} \leqslant t \leqslant t_n$, where $V_q(w, t)$ is the number of packets stored at time t at node w to be transferred further to pole q. The continuity condition, denoted as $A_1(G)$, where $G = \{G_{pq}(t)\}$, $p, q = 1, 2, \ldots, N$, and $t_a \leqslant t \leqslant t_b$ that is a generalization of Eq. (8.43) has the form

$$\frac{dV_q(w, t)}{dt} = G_{w,q}(t) + \sum_{(u, w)\in K_{in}} \Phi_{pq}(u, w, t) - \sum_{(w, v)\in K_{out}} \Phi_{pq}(w, v, t) \tag{8.52}$$

The symbols here are as in Eq. (8.43).

We note that the integral

$$\varepsilon_V[V(\cdot)] = \int_{t_{n-1}}^{t_n} \sum_{q, v} V_q(w, t)\, dt \tag{8.53}$$

has the sense of the total time during which packets remain stored at buffer memories, but only considering the interval (t_{n-1}, t_n). On the other hand, Eq. (8.52) suggests that from $\Phi(\cdot)$, $V(\cdot)$ can be determined. Thus the functional $\varepsilon_V[V(\cdot)]$ is a convenient measure of quality for a time-dependent flow and characterizes the packet transmission time in the network.

In tackling the optimization problem $OP\ \Phi,\ \varepsilon_V \mid A_1(G),\ A_3$, well-known procedures of the solution of deterministic problems in control theory may be applied. The use of these methods, and in particular of the Pontryagin maximum principle, is described in [472].

8.4. Data Teleprocessing Systems Using Random-Access Radio Channel

Introduction. If communication channels in data teleprocessing systems are to be efficient, the ways of their employment should be chosen based on the behaviour of a given network and messages communicated in it.

One of the possible purposes of a computer network is the efficient utilization of geographically distributed computer resources. (As an example of such a network we may refer to the ARPANET network already in operation [414]). This is achieved by providing access of a given terminal not only to its host computer, or through appropriate interfaces to other computers. Individual messages from a sending terminal to a receiving terminal can be communicated mainly in three ways over the network consisting of terminal interface processors and communication links. The first way is telephone circuit switching. The utilization of this method implies the selecting of complete route and providing of communication from the sender to the destination prior to message sending, resulting thus in a substantial delay of the message.

As to delay, a second message switching method is more suitable, whereby the communication is established to the nearest terminal interface processor only. (On the assumption that the message can be stored at each terminal).

A third way, viz. packet switching, may result in a still higher gain in the message delay time. In this method, at the sending terminal the messages are broken down into packets of short length that are transmitted over the communication link as with message switching.

A characteristic feature of this type of computer networks is packet switching.

We may also distinguish another type of communication networks designed to ensure a purposeful access of many sending terminals to a central computer. Here, as is often the case with man-machine systems, each individual sending terminal uses the network a small percentage of time only. Therefore packet switching can be to advantage, whereas for communication one common radio channel for all the sending terminals may be utilized with a somewhat random and statistical separation (random access or statistical multiplexing). This method relies on the central limit theorem in probability theory.

Random access systems feature complete or partial overlap of packets in time, as two or more sending terminals may have ready packets almost simultaneously. Therefore it is necessary to find out whether or not overlaps have occurred and retransmit the packets lost in such a way.

A simple way of overlap identification is characteristic for satellite-repeater systems, as each sending terminal will "hear" its communication after the repeating time has alapsed, that is normally longer than the packet duration. In a ground network, the central controller may send a positive acknowledgement signal to the terminal. In both cases the fact is used that, similar to radio broadcast systems, the message from the center (or central repeater) is received by each terminal. This is one more distinction of random-access systems.

A packet switching computer network is discussed in [461]. In 1970 Abramson proposed a system using random-access principle for a radio link [305], and the ALOHANET project was initiated. After both systems have been commissioned and operated for a long time, reviews have appeared of the first system in [414] and the second in [308]. Lastly, deserving mention are monographs on queueing theory, e.g . [278, 413], which carry all the mathematical formalism required to treat such systems.

Below we consider the features of communication networks using a random-access radio channel [408-411]. The section concludes with two illustrations of practical applications of a random-access channel, the ALOHA system [305, 308], and the experimental system of teleprocessing of electrocardiological information developed by a coauthor of the present book [Szabo, Hungary].

Random-access and "slotted" channel capacity. Now we turn to a general model of a random-access communication network in which one central devise services many users. A configuration, in which each terminal receives communication from the center while it does not receive messages from other terminals, is often called "star-like" system.

Pulses of terminals and the center are data packets of equal duration T, which, besides the message transmitted, also contain the address (identifier) of the destination terminal. In the elementary case, calls may arise at any instant of time, i.e. a given terminal, after the packet has formed, may at once send it on over the channel. The central device, having received the packet correctly, sends a positive acknowledgement signal to the destination terminal. In case of an overlap of two or more packets, there is no reply signal, and appropriate terminals repeat their packets with independent random delays in order that the probability of new overlappings may be reduced.

We will introduce now two important parameters of a random-access channel: transmission coefficient S' equal to the number of correctly received packets per unit time, and load (traffic) of the channel, G', equal to the number of all the packets (correctly received and lost due to overlaps). The corresponding normalized (for the packet duration T) quantities we denote as $S = S'T$ and $G = G'T$. Note that in an ideal time-division system $S = 1$. The maximum value of the transmission coefficient is referred to as the channel capacity.

We calculate the transmission coefficient, depending on the channel loading. The analysis is substantially simplified, if it is assumed that the instants of packets onsets (all packets, i.e. both fresh and repeated) from a Poisson process. This assumption is an approximation as, for one thing, the number of users here is assumed to be infinite, and for the other thing, the repeated packets are regarded

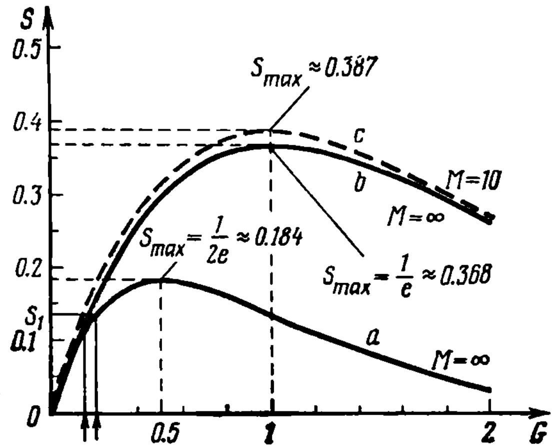

Fig. 8.11. Variation of gain with loading coefficient

as fresh and original. The results of computer-aided simultation indicate a good fit with the Poisson model, if the number of terminals amounts to several tens, and the delay in repeating is on average longer than the packet duration by an order of magnitude.

With this in mind, we obtain that the gain is $S = Gp_0$, where p_0 is the probability of there being no new packets during the time of sending of a given packet plus the preceding time gap T, i.e. during the "dangerous" period equal to $2T$ no additional packets have come into being. The probability p_0 is given by the Poisson distribution

$$p_k(t) = [(\lambda t)^k/k!] \exp(-\lambda t) \tag{8.54}$$

at $\lambda = G$, $t = 2$ and $k = 0$. The required dependence then takes the form

$$S = G \exp(-2G) \tag{8.55}$$

It is presented graphically in Fig. 8.11 (curve a).

In another form of random-access systems, the time axis is broken down into intervals of length T ("slots"), and the transmission of packets is only allowed within these slots. Here two or more packets either overlap completely or do not overlap at all. The dangerous period is now equal to T and the transmission coefficient is

$$S = Ge^{-G} \tag{8.56}$$

(curve b in Fig. 8.11). For brevity, this version hereinafter is called the "slotted" system or "slotted" channel. As is seen in Fig. 8.11, the capacity here is substantially lower than the possible principled limit corresponding to the case of time-division (fixed) multiplexing: $S_{max} = 1/2e$ and $S_{max} = 1/e$ for the two above forms, respectively.

One should not, however, utilize the capacity as the only figure of "merit" as random-access systems normally feature a transmission coefficient below the capacity. Moreover, a maximum of the variations $S = f(G)$ alone is not enough to warrent comparison of various forms of the random-access systems. As is seen in Fig. 8.4, to the point $S = S_1$ there corresponds on the curve b a smaller value of G (i.e. a smaller number of retransmissions and shorter delay), than on the curve a.

We note, that the two forms of the random-access channel discussed above are often referred to in the literature as "pure ALOHA" and "slotted ALOHA" channel, respectively, after the name of the ALOHA system of the University of Hawaii, embodying these principles. We will use these names for brevity.

We now show how the variation of transmission coefficient with load is derived for the "slotted ALOHA" with a finite number of users. It is assumed here that communications of individual terminals (fresh and repeated) are independent events. The loading factor for the mth terminal is $G_m = P$ (the mth terminal transmits a packet in a given slot). The total loading (for one slot is $G = \sum_{m=1}^{M} G_m$, where M refers to the number of terminals in the system.

Similarly, the transmission coefficient for the mth terminal will be $S_m = P$ (a packet of the mth terminal has passed successfully), and $S = \sum_{m=1}^{M} S_m$. Then for S_m we have the relationship

$$S_m = G_m \prod_{i \neq n} (1 - G_i), \quad m = 1,\ 2,\ \ldots,\ M \tag{8.57}$$

If the average activities of the terminals are similar, then, $S_m = S/M$, $G_m/G/M$ and we, by Eq. (8.57), have

$$S = G(1 - G/M)^{M-1} \tag{8.58}$$

For $M \to \infty$ we arrive at Eq. (8.56) derived on the assumption of the Poisson traffic. Eqs. (8.57) and (8.58) give very close values even with small M. The case of $M = 10$ is shown in Fig. 8.11 by a dashed line. Of course, for a finite M the transmission coefficient will be higher than for $M = \infty$.

Transmission factor and average delay of packets. In this section, the analysis of the stationary operation of random-access channels is continued by considering an important index—average delay of packets—in addition to the transmission coefficient. The average delay (D) refers to the mathematical expectation of the time interval from the beginning of a fresh packet to the instant of its a successful reception. The delay time is readily obtainable if the model that is presented earlier is retained, and besides two more assumptions are

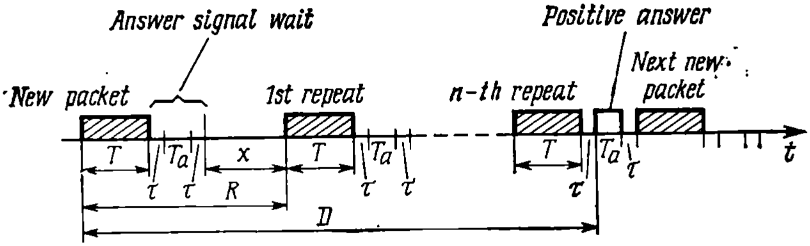

Fig. 8.12. Loading scheme

made: the central controller shapes a reply signal in a negligible time; the reply signal is always received correctly by a terminal.

We denote by R the time span between two adjacent transmissions (repetitions) of a packet. Then

$$R = T + \tau + T_a + \tau + x \tag{8.59}$$

where T is the packet length; τ is the time taken for radio waves to travel from the terminal to the center or back; T_a is the length of the reply signal; x is the value of the random quantity of interval between repetitions.

The average delay time of the packet, is

$$\bar{D} = (G/S - 1)\bar{R} + T + \tau \tag{8.60}$$

where

$$\bar{R} = T + T_a + 2\tau + \bar{x}$$

as in our model the average number of repetitions is equal to $G/S - 1$ [409]. These parameters are shown in Fig. 8.12.

The ratio G/S of Eq. (8.60) in itself is a characteristic of the delay, however, the comparing of various forms of the random-access system calls for a computation of $\bar{D}$, as their values may be different, independent of G/S. An important part here is played by the random quantity x. It might be shown that both its reduction and its growth result in an increase in the packet delays and for each value of S and $\bar{x}_{opt}$ can be found, for which $\bar{D}$ will be minimal.

This optimization is intractable analytically, and the literature provides results for "slotted ALOHA" only [408]. For "pure ALOHA" and two other forms to be discussed below the results are available of the computer-aided studies, which are reproduced in Fig. 8.13. From these curves an estimation and comparison is possible in the important parameter of packet delay.

Dynamic behaviour of random-access radio channels. To study the dynamics of random-access systems, a computer-aided simulation study was undertaken. The experiments have yielded the following results. At the time origin, the system was "empty". In good time

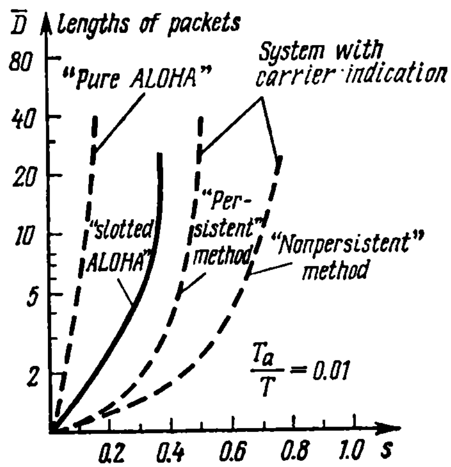

Fig. 8.13. Variation of average packet delay time with transmission coefficient

the system reached an equilibrium state corresponding to a certain (desired) value of the transmission coefficient at a given loading. After a certain time period, statistical fluctuations disturbed the system from an equilibrium condition, the load stepped up, thus decreasing the transmission coefficient, which in turn caused still higher loading, and so forth. The transient process terminated in a transition of the system into another equilibrium state corresponding to channel lockup (minute transmission coefficient at exessively high average packet delay).

This behaviour calls for the introduction of the notion of stability of random-access channels, and also suitable measure of instability, and for a consideration of individual systems.

We will use as a basis the model discussed in [410] and definitions utilized therein. The easiest to analyze is the "slotted ALOHA" with a finite number of terminals M. Each terminal has a memory sufficient to store one packet, and therefore it may be in one of the following two states:

1. The generation of new packets. Here each terminal generates and communicates packets with probability r (in one slot).

2. The blocked state. Here a terminal has in store one packet, whose preceding transmission has been a failure and is to be repeated. The probability that a packet in a given slot be repeated is p.

Let n_t be a random quantity equal to the number of packets stored at a given instant of time t (the number of terminals blocked), and s_t be the frequency of generation of new packets. It is obvious that

$$s_t = (M - n_t)\, r \tag{8.61}$$

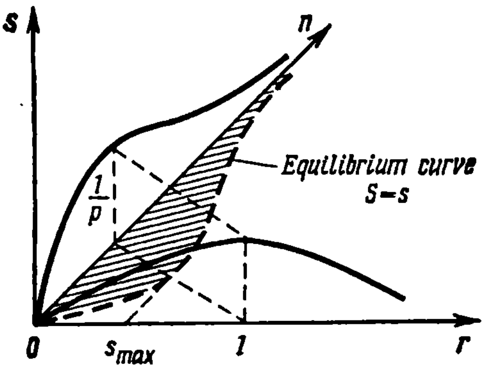

Fig. 8.14. Geometric representation of function S (n, r)

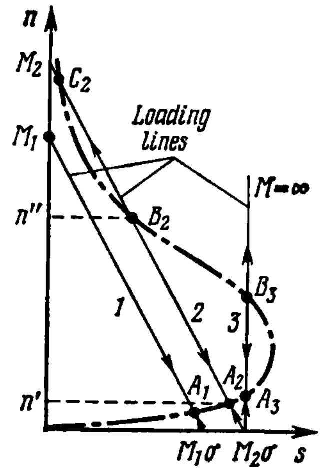

Fig. 8.15. Equilibrium curve and loading line

We will refer to the vector (n_t, s_t) as the vector of channel state. If M and r are time independent, then n_t will be a Markovian process with stationary probabilities of transition.

We denote by S the transmission coefficient (as earlier for the equilibrium state), which now would not be equal to the frequency of generation of new packets: it may be either lower or higher. The transmission coefficient of the channel is conveniently written as the function $n_t = n$ and r; it is equal to the probability that in one slot the same packet is communicated

$$S(n, r) = (1-p)^n (M-n) r (1-r)^{M-n-1} + np(1-p)^{n-1}(1-p)^{M-n} \tag{8.62}$$

Figure 8.14 shows schematically the surface described by Eq. (8.62). For points in the plane (n, r) connected with a dashed line, the equality $S = s$ holds, i.e. the system is in the equilibrium state. This curve may be called the equilibrium curve, on one side of which (shaded in the figure) $S > s$ (the transmission coefficient is higher than the frequency of generating new packets: the difference is due to packets stored), and on the other side $S < s$ (the channel is overloaded, the number of packets transmitted successfully is less than of those emerging).

Given in Fig. 8.15 are the equilibrium curve and several possible positions of so-called load lines [the position of load lines is determined by Eq. (8.61), i.e. parameters M and r]. The arrow heads on the lines indicate the direction of movement of a working point along the load line; in conformity with what has been said above for $s > S$,

the arrows indicate the direction of a higher n number, and for $s < S$ the direction of a lower n number.

The points A_1, A_2, A_3, and C_2 are the points of stable equilibrium and B_2 and B_3 are the points of unstable equilibrium. As proposed in [410], a random-access channel is stable when the load line traverses the equilibrium curve at one point only (case *1* in Fig. 8.15), in other cases the channel is unstable (cases *2* and *3*). Case *3* indicates that a model with infinite terminals is always unstable.

A measure of instability is the average time to the first running out of a "safe zone". ("Safe zone" may refer to the totality of states $n = \{0, 1, \ldots, n\}$ for a given load line). Using the Markov chain theory, the distribution and average value of the first running out [410] can be determined.

When designing random-access system, there are two possibilities: to ensure channel stability (*1* in Fig. 8.15), and to use an unstable channel (*2* in Fig. 8.15). In this case the number of terminals will be higher, but they are served with a measure of reliability (with a certain first running-out time).

By definition of an unstable channel, the stabilization may be performed by way of incorporating a control feature into the system. This implies, for instance, longer repetition times at individual terminals as the channel storage increases and the channel transmission coefficient decreases. The model considered so far may be supplemented by the condition, say, that the terminals repeat their packets in a given slot either with probability p_0, or probability p_1, depending on the instantaneous state of the channel. Suppose that the state of the channel, n_t, is known at each instant of time. In that case, it may be indicated [423] that there exists a solution procedure that maximizes the gain and minimizes the average delay time. According to this procedure, the repetition probability $p = \{p_0, p_1\}$ at each instant of time is determined by the comparison of the state n_t of the channel with a predetermined threshold n^*

$$p = \begin{cases} p_0, & \text{if } n_t < n^* \\ p_1, & \text{if } n_t \geqslant n^* \end{cases} \tag{8.63}$$

The channel state at terminals is known exactly at no time, but it can be estimated from the behaviour of the channel over the preceding time interval. Using such estimates, the control algorithms close to optimal may be derived [423].

Characteristics of carrier sense multiple access systems. When considering earlier the extreme possibilities of random-access systems, we started from the model of a "star-like" system, where terminals require communication with a central device only. But frequently it might be assumed that in such a system, individual terminals may also receive messages from other terminals.* The use of this fact may substantially increase the system efficiency. In fact, if the terminals

are provided with a carrier indicator (common for all the system) a given terminal may, after a next packet has been shaped, transmit it, if the channel is free (no carrier), or delay the transmission to a certain instant in future, if the channel is occupied (carrier is present). It may then be expected that the overlap probability be decreased due to a measure of such kind of order in the system. Nevertheless, by using the name "random-access channel" here also, we imply a significant behaviour of these systems.

The main condition for the use of the carrier indication metod is that the propagation time plus the carrier indication time should be short as compared with the packet duration, otherwise a given terminal will have information about the past state of the channel. (This practically excludes any application of the carrier indication method in satellite communications.)

The carrier indication system behaviour depends on that of the terminal after the channel state indication. Of the variants proposed in [409] two are outlined in short below, "persistent" and "non-persistent" algorithms. When obtaining the results for these systems it has been supposed in [409] that the propagation time and carrier indication time for all the terminals (for all communication routes "center-terminal") are similar. (In practice this condition is a good approximation, as though propagation times may in actual practice be markedly different, they still account for a small portion of the total time.) Other initial assumptions coincide with those used in dealing with the capacity.

The "non-persistent" algorithm in a carrier indication system has the following properties:

1. If a channel is free after a packet has been shaped, then the packet is transmitted at once.
2. If a channel is occupied, the transmission is delayed for a certain (randomly selected) time. Atfer this time has elapsed, a repeated carrier indication occurs.

The transmission coefficient of the channel, depending on the load coefficient, has been derived in [409]. The "dangerous" period here is equal to the sum of propagation and carrier indication times (the value of this time normalized for the packet duration T is denoted further by a), as after this time has elapsed from the beginning of the packet all the other terminals know already that the channel is occupied.

Here it would also pay to divide the time axis into slots with a duration equal now not to the packet duration, but to the value of a. For the "slotted" system with carrier indication in which a "non-persistent" algorithm is utilized, the following relation between S and G is known [409]:

$$S = aG \exp(-aG)/[1 + a - \exp(-aG)] \tag{8.64}$$

It follows from Eq. (8.64) that the transmission coefficient (and the channel capacity) are dependent on the parameter a. With a sufficiently small a, a capacity is attained that by far exceeds that in the ALOHA-type system: at $a = 0.1$, $S_{max} \approx 0.52$, and at $a = 0.01$, $S_{max} \approx 0.81$.

The "persistent" algorithm is the following procedure:

1. If the channel is free after a packet has been formed at a given terminal, the packet is transmitted at once.
2. If the channel is occupied, the terminal waits till the channel is free to start the packet transmission.

The relationship between the transmission and loading coefficients is also derived in [409], it has the form

$$S = G(1 + a - e^{-aG})\, e^{-(1+a)G} / [(1+a)(1 - e^{-aG}) + ae^{-(1+a)G}] \qquad (8.65)$$

Now for $a = 0.1$ we have $S_{max} \approx 0.45$, and for $a = 0.01$ $S_{max} \approx 0.54$. These and other values of S_{max} are much higher than appropriate values for the ALOHA-type channels (see Fig. 8.11). The advantage of the carrier indication systems is also obvious from the comparison of the most important parameter—the average packet delay. Figure 8.12 shows the curves obtained by computer-aided simulation for the two considered forms of carrier indication systems.

Tradeoff with fixed-time and frequency-division systems. It has been noted at the beginning of Section 8.4 that a great many interactive terminals may be served by a random-access system in a more efficient way than by the systems with fixed frequency- and time-divisions.

Here we present some of the quantitative relationships from reference [414].

Initial conditions:

the system consists of M terminals;

the generation of new packets at individual terminals is approximated by a Gaussian process, the generation frequency is then expressed in terms of λ (packets per second), and the packet length is equal to b binary units;

a communication channel with maximum transmission rate V bit/s that may be divided in an arbitrary way;

protective bands between adjacent channels with frequency division may be ignored, and the time taken for synchronization with time division may be ignored.

The comparison is performed according to the average delay of packets which we now designate by δ.

As it cannot be supposed that the packet generation rate is a constant value equal to λ, it is necessary at the terminal that the possibility be provided of storage of several packets waiting to be trans-

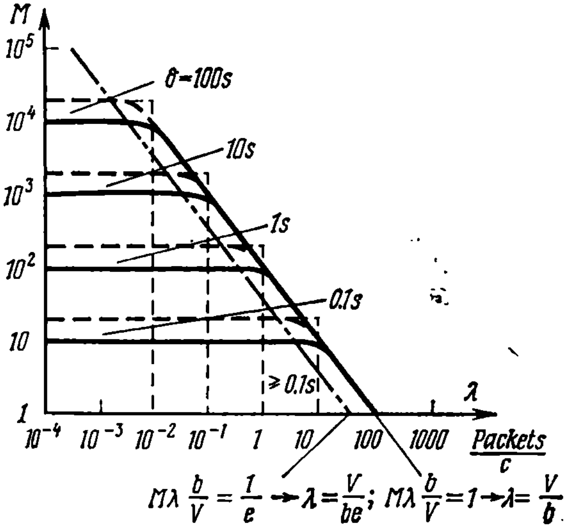

Fig. 8.16. Variation of terminal number with activity of individual terminals: $b = 1000$ bits; $v = 100$ kbit/s; — — — TDM; ——— FDM; —·— "slotted" ALOHA

mitted. This waiting list for low activity of terminals will most likely be empty; however, time intervals will frequently occur in which the generation frequency of new packets will be higher than average and they will spent time waiting to be sent to the communication channel (before they are transmitted).

The average delay of packets with frequency division is equal to the sum of times of waiting and service

$$\delta = M^2\lambda b/2V^2 (1 - M\lambda b/V) + Mb/V \tag{8.66}$$

The first term of the sum is the waiting time obtained for waiting list with an exponental distribution of demands and determinated service time (i.e. for *M/D/1* type systems [413]). The second term relates to the time of passage through the channel.

With time-division multiplexing the average delay time of packets consists of the three components

$$\delta = M^2\lambda b/2V^2 (1 - M\lambda b/V) + Mb/2V + b/V \tag{8.67}$$

The first term in Eq. (8.67) is the waiting time, the second, the average time of waiting for "its" interval in the system with time-division multiplexing, the third is the time of passage through the channel.

Among the various forms of random-access systems we select the "slotted ALOHA" system, as for it the average delay has been defined, depending on the transmission coefficient S (see, e.g. [410]).

Figure 8.16 presents curves to determine the number of terminals

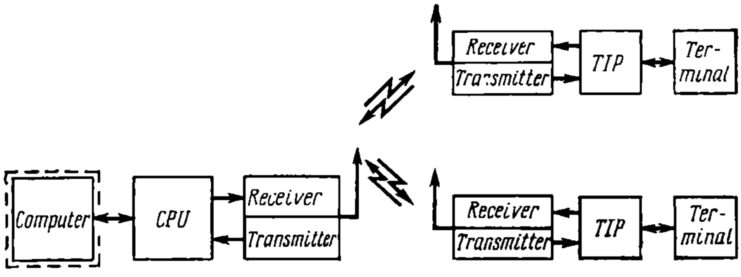

Fig. 8.17. Block diagram of system

that can be serviced by the system, depending on the activity of individual terminals (λ) for a constant δ as a parameter. (For the "slotted ALOHA", the gain factor with a given δ is equal to the constant number $S = M\lambda b/V$, i.e. $M = f(\lambda)$ is a linear dependence.)

It is readily seen in the figure that, for instance, with $\delta = 0.1$ the number of terminals served by a random-access system at small activity of terminals (small λ) is by several orders of magnitude higher that for time- or frequency-division multiplexing. (The last two systems differ from each other much less: the time division enables twice as many terminals to be handled.)

ALOHA system features. A random-access channel was first utilized in the ALOHA computer system. What follows is a short outline of those of the system characteristics which might be of value in designing such type systems.

The ALOHA system was initially based on a truly star-like configuration that was later made more complicated due to repeaters and incorporation into satellite systems. A simplified block diagram of the system center and terminals is presented in Fig. 8.17.

The central processing unit (CPU), which is essentially a small computer, interfaces the system terminals with its central computer. Through CPU the communication through satellites and with other computers is accomplished. The terminals are connected with the system via terminal interface processors (TIP). The tasks of a TIP are as follows: the shaping of packets (this is essentially the addition of an identifier and a code to detect errors); the control of transmission and repetitions; the identification of positive responses and messages arriving from the center to the terminal.

The packet structure is given in Fig. 8.18. Use is made of packets containing either 80 or 40 symbols. The packet duration in second corresponds to the data transmission rate equal to 9 600 binary digits per second (bit/s).

The central processor controls all the message arrivals. It performs the following functions: identification of successfully received packets and generation of positive replies; formation of a queue of messages

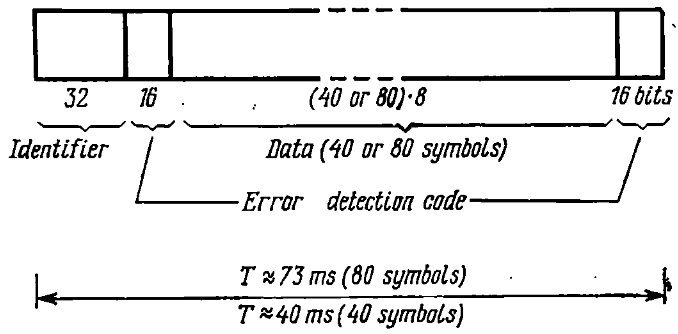

Fig. 8.18. Packet structure

from a computer to terminals in accordance with a certain priority list; provision of superiority of positive replies over computer messages.

The repetition algorithm for overlapping packets is as follows. The packet delay time to the repetition time is a random quantity uniformly distributed within an interval of 0.7 s (around 10 packet durations). After a third unsuccessful repetition, a signal is given to the operator who can perform a new transmission through a manual control. This is equivalent to introducing a rather long delay, thus preventing the channel from locking.

Finally, we would like to estimate, neglecting the duration of average delay and assuming the channel to be stable, how many terminals can be served. In this system it may be roughly supposed that each terminal generates packets with a repetition rate of $S'_i =$ $= 1/60$ packets per second (one packet per minute). Then from the relations

$$S_{max} = MS_i = MS'_i T, \quad S_{max} = 1/2e \tag{8.68}$$

we have for $T \approx 73$ ms

$$M = 1/2eS'_i T \approx [2 \times 2.73 \times (1/60) \times 0.073]^{-1} \approx 154$$

Random-acce s radio channel application in computer processing of ECGs. Systems of teleprocessing of electrocardiological data process with a computer the electrocardiogram of a patient, based on a stored data set and help the doctor in diagnozing. Also, such systems should ensure remote access of several terminals to a central computer.

If the terminals are few and stationary, then telephone links might do for communications. But if they are numerous and the system is to provide for further expansion, then it would be most advantageous to take into account the fact that the total time of examination of a patient is by far longer than required to transmit the electrocardiogram

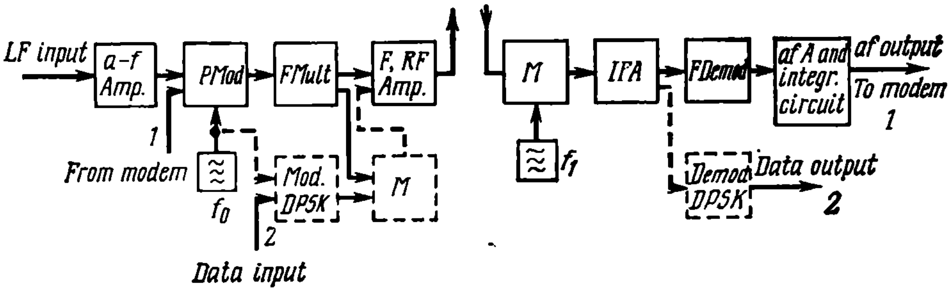

Fig. 8.19. Block diagram of system: PMod—phase modulator; FMult—frequency multiplier; F—filter; M—mixer; FDemod—frequency demodulator a-f A — audio frequency amplifier

and related information. The random-access radio channel offers the following advantages: the system is flexible with a straightforward possibility for expansion; communication link leading costs are much lower than with a telephone network; communication lines are more reliable and may be utilized without limitations.

An experimental system for ECG teleprocessing with the computer EC 1010 has been developed by the Budapest Telecommunication Institute. In this system the so-called medical terminals *MT* and central processor are connected by switched telephone links. Under one of the two operating conditions of *MT*, related information is communicated by using an alphanumerical display and a data modem, and under the other, three analog ECG signals are transmitted from the electrocardiograph by the method of frequency modulation of three subcarriers.

At the Budapest Technological University, a modification of that system is being considered, using “radio terminals” with random access. In the process, characteristic features of information exchange and radio channel used should be taken into account. One of them resides in the fact that the medical terminal transmits two kinds of messages:

accompanying information (code and data of the patient, answers in the form, other results of analysis) makes one 80-character line corresponding to 880 bits with asynchronous operation “start-stop”; thus shaping a packet as in the ALOHA system to obtain a packet length of 1 000 bits;

the transmission time of ECG analog signals is 15 s.

The second feature is the fact that the data transmission rate over the radio channel is governed by the procedure of frequency band division, and by the possibility of using the commercial radiotelephone sets with slight modifications. There are two possibilities: first, to utilize the radio-telephone sets as a telephone channel, i.e. to use the modem output signal for modulation. Thus, a data transmis-

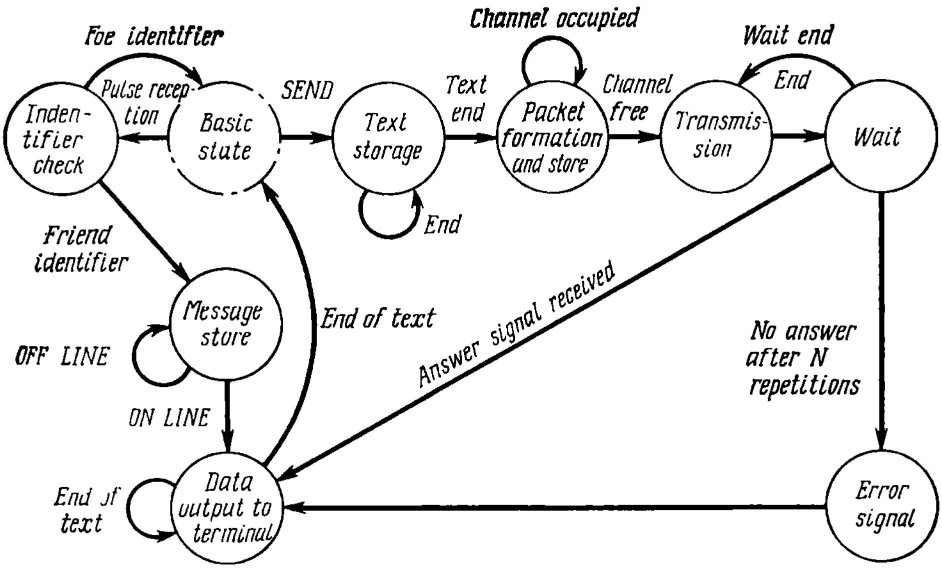

Fig. 8.20. State diagram

sion rate of 1 200 bit/s is readily achieved. For higher rates, more effective modulation techniques in the audio-frequency range should be adopted, thus resulting in a much more complicated system. Second, direct digital modulation of the radio-telephone carrier may be accomplished using major components of the sets. Allowing for a frequency spacing of 25 kHz between channels in the ultrashort wave range, a transmission rate of 10 kbit/s or more is provided with relative ease using, for instance, PDSK. These two possible situations are presented in Fig. 8.19.

With this in mind, the following two approaches might be suggested.

The first one is characterized by the fact that, first, a data transmission rate of 1 200 bit/s is used; second, ECG signals are transmitted in an analog form. The duration of a data packet here is 0.83 s, whereas transmission of a ECG message requires 15 s. Therefore, the data packets have a random access to the central unit based on the carrier indication technique, ECG messages being only communicated if allowed by the center. This permission simultaneously doubles as an inhibit signal for the rest of the terminals to avoid an overlapping of ECG signals.

In the system, the information exchange is controlled by using TIP and CCU in much the same way as in the ALOHA system. The algorithm of the system functioning on the TIP side is shown schematically in the constitution diagram of Fig. 8.20. With a variety

of carrier indication methods available, the "persistent" algorithm was preferred. It is the possibility of conducting easily such experiments that makes a microprocessor for TIP a good choice.

The second approach features the ECG transmission in a digital form and the use of high rates, 9 600 bit/s and more. The duration of the ECG packet is then 2.25 s. (It has been considered here that in the analog-to-digital conversion the count frequency is 300 Hz, and the number of quantization levels is 256. Account has also been taken of the fact that at TIP data can be compressed about five-fold in the algorithm suggested in [476].) In this method, ECG communications may also be transmitted with random access.

Chapter

9

Computer Network Design

9.1. Background

The age of technological revolution saw an unprecedented increase in humanity's requirements for the processing of huge volumes of information.

Advances in electronics and applied mathematics made it possible to satisfy these requirements by way of devising high-capacity facilities for information processing (computer networks) based on geographically distributed means of collecting, processing, and storing information, interconnected by communication lines. It is a common knowledge that earlier computers were designed for solving mathematical problems. A distinction of such a computer usage is a small volume of input and output information and substantial volume of calculations. And even these computers started to be used not only for mathematical calculations, but also for data processing (handling of population census, payroll computations, etc.). First-generation computers were not suitable for large volumes of information. They were not efficient in this application.

Second-generation computers already appeared multipurpose facilities suited both for mathematical calculations and for data processing. At the same time, low-speed input devices used to set a limit to the information processing rate. To remove this disadvantage, multiprogramming systems were created capable of performing several programs simultaneously; during the input of data required for one program the processor may handle another one.

Increased computer capacities necessitated a distribution of computer resources between many users. This task has been solved by using time-division multiplexing systems, in which one computer serves simultaneously scores and hundrends of low-speed users.

The next important development in information processing is associated with the creation of hardware and software required to construct systems with distributed information sources, users, sto-

rage and processing facilities. The solution of this problem called for fundamental integration of information transmitting and processing facilities. Data teleprocessing means and computer networks have been developed [86].

Earlier information processing systems using telegraph hardware to connect users to the computers have been devised in the early 1960s. These systems transmitted with the use of conventional telegraph equipment at relatively low speeds, normally under 110 bit/s. At first, teleprocessing systems were gaining ground at a relatively slow pace, so as late as 1964 the principal way of intercomputer communication of information was via telegraph communication links. But so low transmission rates could not meet the demands of users who were out to devise involved information processing systems.

Another step forward in the development of data communication systems has been the creation of modems enabling binary information to be transmitted over telephone communication links. The earlier prototypes of modems showed a relatively low transmission rate. Later, however, the transmission rate over switched channels has reached 300 bit/s in duplex operation, and 4 800 bit/s in half-duplex operation. Further growth of rates occurs to as high as 1 200 bit/s for duplex, and 9 600 bit/s for half-duplex operating conditions.

Beginning in 1965, special-purpose information processing systems using separate channels have been evolved extensively. Those systems are developed to meet demands of individual organization, and the channel switching facilities are owned by these organizations. The operation of such systems has indicated that computer resources and communication channels are used inadequately, the systems appearing costly and unadaptable to varying conditions. The trend has been revealed for some users to turn to computers with higher capacity for a relatively short time.

All the above has resulted in time-sharing systems being developed, in which most of the users may of their own choice connect, to various information processing facilities. At first, the communication networks were only telephone or telegraph switching communication channels. Later on special-purpose data transmission systems began to be worked out. Early in the 1970s, a radically new switching technique has gained recognition in data transmission. In the method, between sending and receiving ends no direct communication is established, rather the transmission occurs with a data storage at intermediate transit nodes. Each message contains an address part to guide it over the network. At each intermediate node, the communication computer analyzes the addresses of incoming messages or of their parts (packets) and distributes them over buffer facilities where queues are formed for appropriate channels. In such a way the message is being transmitted from one node to another on its way to its destination.

Packet switching systems exhibit a number of useful properties that can in no way be realized in conventional channel switching systems. The advantages of packet switching networks include: versatility of information transmission for users with both large and small volumes of traffic; flexibility of organization of connections with specified points (to enter the network in order to transmit from any point to any other, it suffices to make use of a standard protocol of interfacing with the network); low costs due to all switching and data transmission facilities being used simultaneously by a large number of users; transmission reliability in the network owing to effective error-control means; added reliability due to the possibility of using by-pass routes in case individual hops or nodes go wrong. Packet switching networks are readily modified to cater for changes in information flows, failures and overloads in individual nodes, hence their remarkable reliability and vitality.

The first packet-switching computer network (ARPA) was initiated in the USA early in the 1970s to pool computer resources of geographically dispersed universities and research centers. Using the time-division operation, it enabled the users to reach any computer covered by the network regardless of the location of computer centers. The experience gained with this network allowed for revealing its advantages and disadvantages to give an impetus to further development of a number of advanced networks in many countries.

So, for instance, in France was developed the CYCLADES network, in the United Kingdom the NPL network (National Physical Laboratory), in the USA, late in 1975, a general-purpose data-communication packet-switched network was commissioned (*Telenet*). Similar networks are operating or under construction in Canada (*Datapack*), France (*Transpack*), Japan, Spain and other countries. Transnational networks are coming into being, e.g. the systems European Information Networks (EIN) and *Eoronet* in West Europe. The constructing of joint networks is made possible due to the development of standard procedures of information exchange between terminals and the network. Most of existing packet-switched networks use the network access procedure termed the X25 protocol proposed by the CCITT.

9.2. Data Transmission Techniques in Computer Networks

There are a multitude of ways of combining into a unified system of information collecting, processing and storing facilities relying on various principles of construction of computer networks. The major components of computer networks are: information input and output devices (terminals) which load into the system required information, processing programs, and inquiries and produce proces-

sing results, inquiry answers, etc.; information storage and processing means; information transmission facilities.

Information input-output devices may arbitrarily be subdivided by speed into three classes: low-speed, medium-speed, and high-speed. A typical low-speed input-output device is the terminal with an electric typewriter (teletype). Medium-speed devices are punched-tape devices, displays, and so forth. High input-output speed is required in information transmission between computers, between storages, and so on.

In sophisticated systems, information is stored and processed by multipurpose and special-purpose computers. For the most part multipurpose computers are utilized, whose speed and storage capacity are selected depending on the tasks to be performed. For large data files, the major storages are computer disks and tapes.

Information transmission facilities include three groups of hardware: communication channels, switching systems, and communication channel time-sharing devices.

Communication channels are normally subdivided into three categories: low-speed (with transmission speed up to 200 bit/s), medium-speed (up to 10^4 bit/s), and high-speed (up to millions of bit/s).

Switching systems are broadly classed as follows: channel switching, and message, or packet, switching. The major distinctions of these two techniques are as follows. In channel switching, between the information source and destination switches are used to have a communication channel that is utilized for transmission. In case of message or packet switching, the continuous channel is not constructed and the transmission occurs with intermediate storage at retransmission transit nodes, where messages being transmitted may be stored in memory during the time the channel that leads to the adjacent retransmission node is occupied. The packet-switched system differs from that with message switching in that with long messages these latter are broken down into parts (packets) to be transmitted independently to the destination, where the packets are reassembled into messages and presented to the user. Typically, the packet length ranges from hundreds to thousands of bits.

Packet-switched systems ensure high performance of communication system for relatively short messages typical of an interactive communication with a computer. The channel switching method is effective in transmitting data files.

Communication channel time-sharing systems are meant for better utilization of communication channels through heavier loading with combined information communicated simultaneously for several users. Hardware facilities within this category include: multiplexers combining at the computer input flows from input communication channels to obtain a total high-speed flow; concentrators, remote multiplexers, combining low-speed channels at the input of a high-

speed channel; local distributors for the users responsible for a sequential interfacing with the computer via a common channel of one of the terminals on request or command of the computer; facilities for connection to the common line of several terminals (to organize cluster terminal configurations); communications computers with a variety of functions of managing and controlling the communication at the host computer, a remote location of terminal clustering or switching nodes.

The above options of hardware facilities and configurations of information exchange in computer networks may be used in networks of various types, and their efficiency, cost, reliability and other parameters are controlled by the network's facilities and architecture. Depending on the computer network functions, different configurations may be preferred. Among the most common configurations the following may be given.

Centralized (radial) configuration, in which all the users and auxiliary processor facilities are connected to a host processor concerned with the controlling of operation of the whole of the network. Basic information flows in such a configuration are between the center and users. Between the terminals, information is only communicated via the host processor which here might be referred to as a message switch. An example of such a processor configuration is a time-division multiplexing system, where all the users are interfaced with the host computer.

Hierarchical tree-connected configuration, which has several levels of information processing topped by the master computer. So, for example, in systems controlling large-scale industrial enterprises or branches of industry, small-size computers are used at the lowest level, whose inputs are information from process monitors. These computers can be utilized for the monitoring of process control and shaping of messages about their occurrence (possible troubles, results, etc.) for the next level—coordinating computers. Here data for a number of processes are handled and decisions are made directed at coordinating the remedial actions in case of departures from a predetermined behaviour. The generalized results, that might be used as indications of the performance of a plant, are transferred to the next level concerned with the planning and controlling the operation of a group of similar enterprises, and so forth.

Ring, or loop, connected network is essentially a closed ring communication link passing through all the nodes concerned, which might be computers and terminals. In this configuration, the information transmission management is through the master (host) computer which gives commands to individual nodes to transmit information. The reception occurs in accordance with the address section of the message format directing the message to the appropriate destination.

Such configurations are applicable with relatively short distances

between individual network nodes, e.g. to combine input (output) and information processing facilities in one office. With longer distances between the nodes, the loop configurations appear not cost-effective due to high communications structure cost.

Distributed configuration is the most reliable and viable form. Typically, network nodes are equipped with communication computers connected by separate communication channels with neighbouring nodes. To the communication computers are connected information processing computers and the terminals (through appropriate interfaces). Between the nodes information is normally communicated using the channel, message, or packet switching.

The distributed systems have no center supervising the network operation, though at times the network may have a center pooling the information on the network condition, which is used to work out recommendations as to routing the messages.

9.3. Suboptimal Methods of Network Design

Introductory remarks. For the most part, the effort of designing computer networks is essentially a combinatorial optimization problem of high dimensionality.

In recent years, within the framework of mathematical programming a variety of effective methods of searching for optimal solutions for various classes of problems have been devised. The most interesting results have been obtained for problems solved by linear programming techniques. Solution procedures have been worked out for certain classes of nonlinear and integer-valued problems [145].

At the same time, quite a number of complex problems of higher dimensionality defy solution by the devised linear programming techniques owing to the fact that these either do not reduce to classes of problems with known procedures for arriving at an optimal solution, or do not satisfy purely practical constraints imposed by a limited memory volume or speed of the computers utilized.

To handle these problems, suboptimal methods would be of value, which do not by all means give an optimal result, but provide a "good" solution meeting the problem constraints. In so doing, that "good" solution may progressively improve and come nearer to an optimum with more computer time expended.

Problem formulation. In the general case the problem of suboptimal programming can be written as follows.

Let G be a set of all allowable solutions. We will suppose that for each feasible solution $g \in G$ and efficiency criterion $H(g)$ can be determined, and a solution g_i is regarded as better than a solution g_j if $H(g_i) > H(g_j)$.

The solution g^* is optimal if the following condition is satisfied:

$$H(g^*) \geqslant H(g) \quad \forall g \in G \tag{9.1}$$

We now introduce the notion of the suboptimal algorithm. To this end, certain auxiliary quantities will be defined below.

Let it be so that to reach a level of the objective function H^*, using the algorithm α_i, a computer time t_i is required, and to arrive at the same level by the random equiprobable sampling method, the time t_0.

The algorithm α_i will then be referred to as suboptimal, if the relative complexity criterion is $\mu_{i0} = t_i/t_0 \ll 1$. In other words, the algorithm α_i will be said to be suboptimal if it leads to a substantial reduction in computer time taken to attain a certain level of the objective function as compared with the algorithm of the random equiprobable sampling.

The algorithm α_i will be called asymptotically optimal, if for a limited increase in computer time it results in an optimal solution, i.e. if

$$\lim_{t \to T}[H(g^*) - H(\hat{g})] = 0 \tag{9.2}$$

where t is the computer time for the algorithm α_i; T is a sufficiently large, but finite, quantity; $\hat{g}$ is the best solution for the time t.

It is to be noted that exhaustion and random equiprobable sampling procedures of seeking feasible solutions of the set G provide, in principle, an optimal solution; however, their practical realization for higher-dimensionality problems is exceedingly cumbersome and not feasible with practical limitations on the computer memory and time.

We will focus on the problem of this very type, i.e. we will assume that for the problem at hand for one or another reason, conventional techniques of mathematical programming yielding optimal solution cannot be applied; and that the dimensionality of the set G is such that the complete exhaustion is not possible.

Under these conditions, especial significance is attached to hypotheses or heuristic assumptions about the most promising directions of searching which might yield good or optimal solutions. These hypotheses may be shaped by the investigator directly in the course of the solving of a problem or prearranged in the form of appropriate rules (heuristics) realized by using algorithms, these rules govern the computer search procedure.

In the first case the problem is solved in an interactive mode for which the investigator shapes the hypotheses as to promising directions of seeking solutions, and the computer is utilized to implement the search procedure and process the results obtained.

The procedure is iterative and as it proceeds the investigator tests the hypotheses proposed by estimating the results of preliminary recommendations. In the second case the search procedure is carri-

ed out by the computer, with the investigator partaking only in the selecting of heuristics used in the search.

We now focus on certain concrete procedures.

Heuristic branching. This section deals with the procedure of the man-computer interactive search, in which the investigator draws on his knowledge of the problem at hand to seek the principal direction of search, with the computer performing the searching, processing the results and providing information for the investigator so that he can estimate the hypotheses realized and improve them.

The heuristic branching method [233, 468], in essence, consists in the following. Let there be at a certain stage several possible search directions (search tree branches). Then the investigator's task is to indicate the most promising direction, in his opinion, or arrange all the possible directions by the heuristic estimation of their promise. Next, the computer generates feasible solutions for all possible directions, but the greatest attention is paid to those directions which the investigator regards as the most promising. The investigator refines his assumptions from the computer results and the procedure recurs. Here the existence of realizable algorithms is assumed for the generating of feasible solutions belonging to the predetermined search directions.

The selecting of search direction is equivalent to the separating of a subset of a set of feasible solutions G, and with this in mind, we will describe the search procedure in terms of sets G.

The procedure of heuristic branching is as follows. Let at the kth step of the search the set of all feasible solutions G^k be divided into r_k subsets G_i^k, $i = 1, 2, \ldots, r_k$. All the sets G_i^k can be divided into three classes:

subsets, for which the best solution $g_i^* \in G_i^k$ is known. We will denote the class of these subsets by Q_k^*;

subsets, for which the upper limit of the efficiency criterion $H(G_i^k)$ is known, i.e. for which it can be shown that the best solution g_i^* (that is not found yet) has the efficiency criterion $H(g_i^*) \leqslant H(G_i^k)$. We will refer to the class of these subsets as $\tilde{Q}_k$;

subsets, for which there are no rigorous estimates of the upper value of the efficiency criterion. We will denote these subsets as Q_k.

We designate the best solution (record) found in the preceding stages by $\hat{g}$ and the heuristic estimate of the promise of the set G_i^k by $\varphi(G_i^k)$. The heuristic branching procedure may then be described by the following algorithm:

0. Set initial values: $k = 1$, $G^k = G$.
1. Divide the set G^k into subsets $G^k{}_i$, $i = 1, 2, \ldots, r_k$.
2. Find for all the subsets of Q_k^* the best values of g_k^*.
3. Determine for all the subsets of $\tilde{Q}_k$ the value of $H(G_i^k)$.

4. In the set G^k truncate those subsets of G_k^* or $\hat{G}_k$, where solutions are absent which are better than the peak attained.

5. Arrange the remaining subsets r'_k in order of decreasing heuristic estimate $\varphi(G_i^k)$. Let the ordered sequence be

$$G_{i_1}^k, \ldots, G_{i_{r'_k}}^k, \quad r'_k \leqslant r_k$$

6. Using any procedure (heuristic search, random selection, directed exhaustion, etc.), find for each subset $G_{i_j}^k \not\subset Q_k^* m(j)$ different solutions at $m(j_k) \geqslant m(j_l)$, if $j_k < j_l$. For the solutions obtained determine local records $\hat{g}_{ij}$ and, if necessary, refine the total record $\hat{g}$.

7. Assess the feasibility of continuing the search and with a positive solution increase k by 1.

8. Arrange subsets G_{ij}^k in order of decreasing $H(\hat{g}_{ij})$, and the first subset of the ordered sequence that does not belong to Q_k^*, divide into r_k parts. Return to step 2. For discrete problems, the above procedure is asymptotically optimal. To prove it, suffice it to show that in truncating the set G^k (item 4 of the algorithm) the subset cannot be discarded that contains the optimal solution. This is really the case as only those subsets are truncated which contain no solutions better than those obtained. Therefore with an arbitrarily poor search procedure with a finite number of steps, at each of which one of the subsets is divided, an optimal solution will be found.

Though the heuristic branching procedure is asymptotically optimal, its efficiency, however, is strongly dependent on the validity of the hypotheses put forward by the investigator. If they are true, then the procedure enables the search time to be markedly reduced as compared with the time required for exhaustion or random equiprobable selection. This improvement is unattainable in the absence of true hypotheses.

If we ignore the requirements of asymptotic optimality, then the seeking of suboptimal solutions may be simplified and expedited truncating subsets that do not show promise. These include, for instance, subsets which during a sufficiently long period of time of search occupy the last places in the ordered sequence, if the number of the considered elements of these subsets accounts for a rather high proportion of the total number of elements.

The method of fuzzy heuristics was used in ALGOL to seek solutions to the problem of distribution of complex algorithms in the computer network [7, 207], and the findings pointed to high efficiency of the method for the above examples.

Adaptive heuristic search. In the general case, a feasible solution $\mathbf{g} = (g_1, \ldots, g_n)$ may be sought in M stages, in each of which a certain subset of coordinates $\mathbf{g}$ is determined.

We denote the set of coordinate numbers for the jth stage by $S_j = \{j_1, \ldots, j_{m_j}\}$, $j = 1, \ldots, M$ and will assume that $\bigcup_j S_j = \{1, \ldots, n\} \bigcap_j S_j = \varphi$. And the subset of coordinates to be determined in the jth stage we will designate as $\bar{g}_j = \{g_{j_1}, \ldots, g_jmj\}$ and refer to as the jth local solution.

The step-by-step solution search may be exemplified by the successive determination of coordinates of the vector g, when $M = n$. The arguments that follow refer to this very case. We will suppose that in the jth stage of solving the values of g_j may be chosen from the finite set $\bar{G}_j$, containing N_j various local solutions, $\bar{G}_j = \{g_1^j, \ldots, g_N^j\}$. In the general case the set $\bar{G}_j$ may be sensitive to the solution chosen in the preceding stages, i.e. to $g_1, \ldots, g_{j-1}$.

Further, let in selecting the local solutions in all the stages make use of a series of z computable hypotheses (heuristics), determined by the vectors $\Gamma_s = (C_{s1}, \ldots, C_{sN})$, $s = 1, \ldots, z$. In the jth stage of searching the values of components C_{si}^j of the vector $\Gamma_s(j) = (C_{s1}^j, \ldots, C_{sN_j}^j)$ determine the feasibility of selection in this stage of the ith solution, where $i = 1, \ldots, N_j$.

Now let us assume that the procedure of calculation of values of C_{si}^j is such that $C_{si_1}^j > C_{si_2}^j$, if under the hypothesis Γ_s the solution $g_{i_1}^j$ in the jth stage is more advisable than the solution $g_{i_2}^j$. We will refer to the components C_{si}^j as local criteria.

Individual hypotheses Γ_s may be more or less effective, depending on concrete conditions of the problem at hand. In addition, the effectiveness of the hypotheses may be different in various search stages. The adaptive procedure is aimed at the shaping of effective compositions of hypotheses for the various stages of solution searching resulting in a speedy search of suboptimal solutions.

The general approach applicable to the construction of an adaptive procedure for obtaining the hypotheses composition consists in varying the contribution of various hypotheses to the general procedure of solution selection, depending on the validity of recommendations obtained under appropriate hypotheses at preceding steps of search. This approach may, for example, be implemented as follows.

We will compute the local criteria C_{si}^j, corresponding to the sth hypothesis so that:

$0 \leqslant C_{si}^j \leqslant 1$, if under the sth hypothesis it is expedient to choose the ith solution in the jth stage;

$C_{si}^j < 0$, if the ith solution is allowed, but not recommended under the ith hypothesis;

$C_{si}^j = C^* > 1$, if it is allowed to choose the ith solution under the conditions of the problem.

So, for instance, if the initial calculated values of the criterion C_{si}^{j} decrease with increasing advisability of selecting the ith solutions, then to compute the local criteria C_{si}^{j} the following transformations may be carried out:

$$C_{si}^{j} = (\hat{C}_{s}^{j} - \hat{C}_{si}^{j})/C_{s}^{j}, \text{ if the } i\text{th solution is feasible;}$$

$$C_{si}^{j} = C^{*}, \text{ if the } i\text{th solution is not feasible;}$$

$\widetilde{C}_{s}^{j} = \frac{1}{n} \sum_{i=1}^{N_j} C_{si}^{j}$ is the average value of the sth local criterion in the jth stage.

The generalized local criterion C_{i}^{j} for the ith solution in the jth stage can be then defined as the weighted sum of z special criteria with weighting factor W_{s}^{j}, $s = 1, \ldots, z$, for $|W_{s}^{j}| \leqslant 1$ and $W_{s}^{j} \neq 0$:

$$C_{i}^{j} = \frac{1}{z_{i}^{j}} \sum_{s=1} W_{s}^{j} C_{si}^{j} \text{ for } C_{si}^{j} \neq C^{*} \text{ at no value of } s;$$

$$C_{i}^{j} = C^{*} \text{ for } C_{si}^{j} = C^{*} \text{ at least at one value of } s.$$

In the jth stage, the solution is selected in accordance with the obtained generalized local criteria so that those feasible solutions are preferred for which C_{i}^{j} is the greatest. Such a procedure might rely on fuzzy heuristics to be considered in the following section.

After a certain number (say, N) of solutions have been found, we proceed to estimate the effectiveness of various heuristics utilized in the search process. The estimating may be performed by the averaged contribution of a given heuristic to the betterment of the record obtained earlier. If the recommendations derived following a given heuristic in certain stages made for improvement of the record, then the contribution of this heuristic in these stages should be increased. Should the recommendations of the heuristic in question not coincide with the selected solutions improving the record, the contribution of the heuristic to appropriate stages should be dwarfed.

One of the special cases is the procedure at $N = 1$ when a modification of weighting factor occurs after each solution. In this case after M stages of search have been performed, the solution obtained is compared with the earlier record, and a modification of weight coefficients is carried out. With a solution improving the record, a "positive" modification of the weight vector occurs, where the weight coefficients W_{s}^{j} are increased by a value ΔW_1 for those values of s and j, for which C_{si}^{j} were positive for the i that yielded a good result, or decreased by ΔW_1, if C_{si}^{j} were negative.

A similar correction by ΔW_2, but with the opposite sign can also be made after a solution that does not improve the record. We will call

this modification "negative". In this case the modification makes a transition possible to another search area corresponding to a new composition of the local criteria used.

The above procedure was implemented in FORTRAN to attack the travelling-salesman problem. In this program two heuristic hypotheses on the selection of search direction were utilized. Under a first hypothesis, local criteria were the higher, the shorter the distance from the point were the salesman is, to the ith point under consideration. A second criterion was such that its values of C^j_{si} were the higher, the smaller would be the sum of all the distances to the points passed in the transition of the salesman from a current point to the ith point.

The separate testing of these hypotheses indicated that the first is the more effective, i.e. ensures a faster approaching of the solution to an optimal value. It was shown in the examination of the variation of weighting coefficients of the two hypotheses in the course of the program implementation using a positive modification, that under the first hypothesis the coefficients grow faster, therefore in a time the first hypothesis began to exert major influence on the selecting of solutions.

Fuzzy heuristics. The heuristics being not rigid constraints, in the general case the area truncated by them may contain an optimal problem solution beyond the search area, i.e. the search will not be asymptotically optimal. In a number of cases, heuristic search procedures lead to a unique solution. Under these conditions the effectiveness of heuristics is strongly dependent on a specific problem, and the solution obtained may be appreciably different from optimum.

The objective of this section is to describe the asymptotically optimal procedure of heuristic solution search for discrete problems, depending upon the substituting of less rigorous rules for unique heuristic rules fuzzied by recommendations as to search in the directions determined by heuristics [235]. To implement this approach, use is made of procedures of controlled broadening (fuzziation) of the search area determined by heuristics, the degree of fuzziation being different, depending on limitations on the search time and other factors.

Our further reasoning will require a measure of closeness or separation between local solutions. This separation may be determined in a number of ways, but we will only be interested in a strategy of its determination with a measure of separation taking into account the degree of correspondence of local solutions to a certain heuristic hypothesis.

We will begin with an example. Let the travelling-salesman problem in selecting a local solution use a heuristic of transition to the neighbouring point. Then the separation between local solutions will be the less, the smaller the difference in spacing between the

appropriate points and the current point. In other words, if the distances from a point k, where our salesman is at the moment, to points i and j are r_{ki} and r_{kj}, respectively, then under the hypothesis the distance between the solutions i and j will be $d(i, j) = |r_{ki} - r_{kj}|$.

In the general case, if the advisability of choosing the solutions g_i^j and g_k^j in the jth stage under hypothesis Γ_s is determined by the local criteria C_{si}^j and C_{sk}^j, respectively, then the separation between the local solutions g_i^j and g_k^j will be determined under hypothesis Γ_s as $d_s(g_i^j, g_k^j) = |C_{si}^i - C_{sk}^j|$.

The local solution $g_{i_0}^j$ is said to be the closest under hypothesis Γ_s if for any $l \neq 0$ the inequality $C_{si_0}^j \geqslant C_{si_l}^j$ is satisfied.

Also under hypothesis Γ_s, the Δ-neighbourhood of a local solution $g_{i_0}^j$ will refer to such a subset G_{js}^{Δ} of the set of local solutions G_j that for all the elements of the subset the condition $d_s(g_{i_0}^j, g_{i_l}^j) \leqslant \Delta$ is obeyed.

Heuristic search area may be broadened by two procedures, determinate and probabilistic. In each stage of the first strategy the value of Δ is selected so that the Δ-neighbourhood would contain a given number of points. The procedure for searching suboptimal feasible solutions consists in sequential exhausting all the feasible solutions for which local solutions are chosen within Δ-neighbourhoods.

With determinated broadening the search calls for the sequential analysis of feasible solutions to be properly organized, and the information on the variants considered to be stored. In some instances, these conditions may appear unacceptable owing to excessive storage requirements. These requirements can be made much easier by using the procedure of random selection of local solutions within Δ-neighbourhoods.

Probabilistic broadening. With probabilistic broadening, each local solution g_i^j is assigned a probabilistic measure $p(g_i^j)$, such that

1. $0 \leqslant p(g_i^j) \leqslant 1$
2. $p(g_i^j) \geqslant p(g_k^j)$, if $d(g_{i_0}^j, g_i^j) \leqslant d(g_{i_0}^j, g_k^j)$, $g_i^j, g_k^i \in G_j$
3. $\sum\limits_{g_i^j \in G_j} p(g_i^j) = 1$

With such a procedure, in each search stage the local discrete distribution is constructed, and the local solution is sought in a random manner according to this distribution. To broaden or narrow the area of selecting the local solutions (which is equivalent to the broadening or narrowing of Δ-neighbourhoods), the local distributions

may be transformed. Consider by way of example some possible transformations.

Let in a certain search stage the local distribution has the form

$$\mathbf{p} = (p_1, \ldots, p_N)$$

Then to change the Δ-neighbourhood, transformation may be used, say, of the following types:

$$1. \; p_i' = \begin{cases} p_i/p_\Sigma, & \text{if} \quad p_i \geqslant p_0 \\ 0, & \text{if} \quad p_i < p_0 \end{cases}$$

where p_0 is the threshold value, $0 \leqslant p_0 \leqslant p_{\max}$; $p_{\max}$ is the maximum value, for p_i, $i = 1, 2, \ldots, N$, $p_\Sigma = \sum_{p_i \geqslant p_0} p_i$.

$$2. \; p_i' = p_i^L \Big/ \sum_{i=1}^{N} p_i^L, \; L \geqslant 0$$

In the first transformation, the Δ-neighbourhood narrows as p_0 grows, and it may converge to a unique solution for $p_0 = p_{\max}$. In the second case, the Δ-neighbourhood narrows with increasing L and broadens with decreasing L, including at $L = 0$ all the local solutions with equal probabilities of selection. The searching may occur with sequential decreasing of L, i.e. sequential broadening of Δ-neighbourhood.

The solution search technique using fuzzy heuristic in discrete problems with sequential broadening of Δ-neighbourhood is asymptotically optimal. In fact, the broadening of Δ-neighbourhood diminishes the constraints due to heuristic hypotheses, and the search area is bound to contain an optimal solution. With discrete variables, the number of feasible solutions is finite, and hence an optimal solution will be found in a finite number of steps.

The effectiveness of searching, i.e. the pace at which the optimal solution is approached, is notably dependent on the heuristic hypothesis selected; however, even with the most unfavourable selection the fuzzy heuristic method guarantees the optimal solution found in a finite number of steps. Conventional procedures of heuristic searching based on rigidly limited search area do not provide such a guarantee.

The fuzzy heuristic method has been realized in one of the forms of the above-mentioned program of solving the travelling-salesman problem. Given below are some of the results obtained in experimental investigation of the program [236] as applied to the problem described in [385].

To change the degree of fuzziness of heuristic procedure according to generalized local criteria, the probabilistic broadening of the second type with variable L was used. With larger L the

heuristic rules are adopted to limit sufficiently the search zone, and in that case similar results are repeatedly obtained. With larger fuzziness of heuristics (lower L), in the initial search region the progress is much slower; however, in a time the program yields better results, as compared to larger L. For each given search time interval T there exists an optimal value of L giving best results.

Suboptimal dynamic programming. The philosophy of dynamic programming consists in using partial sequential optimization within extending search areas. The search area is thought of as a series of progressively extending areas enclosing each other. This replaces the process of seeking the optimum of function of n variables by an n-step process of searching the extremum of a function of one variable. Though the principles underlying the dynamic programming are rather simple, its practical implementation is complicated. This is associated not only with the absence of sufficiently general algorithms used to handle various classes of problems, but also with the exponential growth of storage volume and the amount of calculations with problems of higher dimensionality. The complexity of solution is strongly influenced by the number of variables and the number of values these latter may take on.

In addition, as the solving progresses the computer memory is to store a great many intermediate results, which is also a limiting factor in the use of dynamic programming to attack real practical problems of higher dimensionality [184, 477].

The objective of this section is to seek ways of curtailing the amount of calculations and storage required through the transition from the optimal solution search to a search for suboptimal solutions approaching progressively to the optimal one with longer computer times. We will discuss the general approach, using for illustration worked problems, in particular the travelling-salesman problem.

Let $n+1$ towns be given with the separation between them determined by the matrix $(C = C_{ij})$, $i, j = 0, \ldots, n$. The shortest closed route is sought that passes through each town but once.

Now, let before the beginning of the kth stage of route selection, a set of points $i_0, i_1, \ldots, i_{k-1}$ be determined and a suboptimal route be indicated for passing from point i_0 to point i_{k-1}, say, $S_{k-1}(i_0, \ldots, i_{k-1})$, $1 \leqslant k \leqslant n$. And further, let we have a procedure of selecting the kth point to be included into the route at kth step. (Some of the approaches to constructing this procedure will be discussed below.)

Such a procedure of extending the suboptimal route S_k may be derived as follows. We obtain

$$\begin{aligned} f_k(i_0, \ldots, i_k) = \min\{&[C_{i_0 i_k} + C_{i_k i_1} + z(i_1, i_{k-1})], \\ &[z(i_0, i_1) + C_{i_1 i_k} + C_{i_k i_2} + z(i_2, i_{k-1})], \ldots, \\ &[z(i_0, i_{k-1}) + C_{i_{k-1} i_k}]\} \end{aligned}$$

where $z(i_j, i_l)$ is the length of suboptimal route $S_{k-1}(i_0, \ldots, i_{k-1})$ from point i_j to i_l, for $j < l$. Let a minimal value $f_k(i_0, \ldots, i_k)$ correspond to the commutation $S_k(i_0, \ldots, i_j, i_k, i_{j-1}, \ldots, i_{k-1})$. Then this very commutation is the result of searching for a suboptimal routing in the kth stage.

Such a strategy retains "good" route sections determined in the earlier stages, in particular between points i_0 and i_j and between points i_{j+1} and i_{k-1}. Here the resultant route may be thought of as suboptimal.

To the route found, a method may be applied of local optimization that is based upon cutting the route length through some commutations of route points. It is easily shown that with such an approach the amount of computations is governed by a power function of the form $c(n^2 + n)$, where c is a constant. This essentially requires no additional storage capacity for intermediate results.

Next we consider some selection strategies for new points to be incorporated into the sequence, ensuring the construction of a class of suboptimal solutions. Let certain heuristics be given, which result in "good" candidates for inclusion into the sequence with due account for some of the local criteria. For the travelling-salesman problem these criteria might be exemplified, in particular, by the distance from the last point of the route constructed to each point not included in the route. As a heuristic hypothesis, the recommendation to transfer to an adjacent point may be used. To arrive at a candidate to be included in the sequence, in each stage the computation is required to be proportional to the number of points uncovered.

Still another example is the heuristic controlling the selection of the point i_k and satisfying the condition

$$f_k(i_0, \ldots, i_k) = \min_{i_k \in A} \min_{i_j} [z(i_0, i_j) + C_{i_j i_k} + C_{i_k i_{j+1}} + z(i_{j+1}, i_{k-1})], \quad k = 1, \ldots, n$$

where A is the set of uncovered points; i_j is a position in the sequence constructed, after which the point under consideration, i_k, lies, $1 \leqslant j \leqslant k-1$.

In this case the volume of computations is proportional to the product of the number of positions in the sequence, where the point under consideration may lie, and the number of possible estimated, i.e., uncovered, points. In the kth stage the amount of computations is equal to $Ck(n - k + 1)$. The total amount of computations here amounts to

$$\sum_{k=1}^{n} C(nk - k^2 + k) = C/6\,(n^3 + 3n^2 + 2n)$$

where C is constant.

Using the fuzzy heuristic method in selecting a next candidate to be incorporated into the route being constructed, a class of suboptimal solutions may be generated and the optimal solution may be sought step-by-step. The major advantage of such an approach is that no additional memory is required to store intermediate results.

The method of suboptimal dynamic programming has been applied to one of the forms of the earlier-mentioned travelling-salesman problem. In that case a next point to be included in the route being constructed was chosen by fuzzy heuristics method using the heuristic hypotheses discussed above. After the next point has been chosen, a local optimum in location of the selected point on the suboptimal route being constructed was determined. In other respects the program was similar to the above algorithm.

The results indicated that the suboptimal dynamic programming may be a rather effective tool in solving the optimization problems of combinatorial type. The quantitative characteristics are discussed in the following section.

9.4. Some Experimental Results

Given below is a summary of comparative results characterizing the effectiveness of some of the above algorithms in handling combinatorial optimization problems. As a test problem for a tradeoff of algorithms, use was made of the travelling-salesman problem at $n = 25$, whose inputs are described in reference [385]. The algorithms were compared by the complexity test defined at the beginning of Section 9.3, and by the effectiveness test to be defined below.

With random equiprobable solution selection (algorithm α_0), four experiments under various conditions with 3 240 million computer operations (3 240 MCO) each gave no improvement over the case with $H^* = 3\,463$. But using the fuzzy heuristic technique and hypothesis of preference of nearest points (algorithm α_1) after the first step with 1.5 MCO gave in five experiments the results meeting the condition $H^* \leqslant 2\,300$. The suboptimal dynamic programming (algorithm α_2) after the first step that also required around 1.5 MCO, produced $H^* = 1\,961$.

Under these conditions, there is no calculating the exact value of μ_{0i}, however, it may be written that for both algorithms $\mu_{0i} > 3\,240/1.5 = 2\,160$, which suggests that these algorithms may be classed with the suboptimals.

The tradeoff of algorithms α_i and α_j may also use the effectiveness test

$$\eta_{ij}(s^*) = \begin{cases} H_i(s^*)/H_j(s^*) & \text{for maximization problems} \\ H_j(s^*)/H_i(s^*) & \text{for minimization problems} \end{cases}$$

where H_i (s^*) is the value of the objective function attainable by the algorithm α_i with s^* computer operations. For the minimization problems, this test determines the ratio of the value of objective function for the algorithm α_i to that with the algorithm α_j, the computer times being equal.

The findings characterizing the performance of the algorithms α_0, α_1, and α_2 are summarized in Table 9.1.

TABLE 9.1

s^* (MCO)	1.5	10	100	500	1000	3000
H_0 (s^*)	4386	4060	3655	3589	3589	3589
H_1 (s^*)	2097	1959	1933	1856	—	—
H_2 (s^*)	1961	1803	1768	—	—	—

Table 9.2 lists the values of effectiveness test for algorithms α_1 and α_2 as compared with those obtained using the equiprobable random selection.

These data attest that for the problem in hand, the algorithms α_1 and α_2 are by far more effective than α_0. These give a value of the

TABLA 9.2

s^* (MCO)	1.5	10	100
η_{10} (s^*)	2.09	2.07	1.89
η_{20} (s^*)	2.24	2.25	2.06
η_{21} (s^*)	1.07	1.09	1.09

objective function that is about twice as good as that obtained by the equiprobable random selection algorithm, the computer times being equal. The algorithm combining the fuzzy heuristic technique with the suboptimal dynamic programming (α_2) is more effective than the fuzzy heuristic algorithm (α_1). To achieve similar results by using the above suboptimal algorithms takes by three order of magnitude less search time.

Bibliography

1. Аболиц А. И., "Энергетические соотношения при передаче сигналов с частотным разделением через нелинейный ретранслятор", *Электросвязь*, 1967, **3**, 1-9.
2. Агеев Д. В., "Основы теории линейной селекции", *Научно-технический сборник ЛЭИС*, 1935, **10**.
3. Агеев Д. В., Бабанов Ю. Н., "Радиоприем при перекрещивающихся частотных спектрах полезного и мешающего АМ сигналов и флуктуационных шумах", *Электросвязь*, 1965, **2**, 1-8.
4. Алексеев А. И., Шереметьев А. Т., Тузов Г. И., Глазов Б. И., *Теория и применение псевдослучайных сигналов*, Наука, Москва, 1969.
5. Альперт А. Я., *Распространение радиоволн и ионосфера*, Изд-во АН СССР, 1960.
6. Альтман Л., "Приборы с зарядовой связью в ЗУ и аналоговых процессорах", *Электроника*, 1974, **16**, 25-37.
   Altman, L., "Charge-coupled Devices Move in on Memories and Analog Signal Processing", *Electronics*, 1974, 8, 91-101.
7. Амбаров Е. В., "Подходы к векторной оптимизации на вычислительной сети", в кн.: *Кодирование и передача информации в вычислительных сетях*, С. И. Самойленко (ред.), Сов. радио, Москва, 1977.
8. Андронов А. А., Витт А. А., Понтрягин Л. С., "О статистическом рассмотрении динамических систем", *Журнал экспериментальной и теоретической физики*, 1933, т. 3, **3**, 165-180.
9. Андронов И. С., Финк Л. М., *Передача дискретных сообщений по параллельным каналам*, Сов. радио, Москва, 1971.
10. Ач Э., "Некоторые вопросы построения адресно-кодовой системы связи", *Радиотехника*, 1968, **6**, 85-89.
11. Бабанов Ю. Н., "Повышение эффективности одного способа улучшения избирательности радиоприемных устройств", *Электросвязь*, 1963, **11**, 64-68.
12. Бабкин В. Ф., Крюков А. В., Штарьков Ю. М., *Сжатие данных. Аппаратура для космических исследований*, Наука, Москва, 1972.
13. Бадалов А. Л., "О некоторых принципах распределения радиочастотного спектра и составления планов для различных радиослужб", *Электросвязь*, 1965, **3**, 1-12.
14. Бадалов А. Л., Пчелкин В. Ф., "Электромагнитная совместимость радиоаппаратуры и распределение радиорасчетного спектра", *Радиотехника*, 1967, **10**, 1-6.
15. Басс Ф. Г., Фукс И. М., *Рассеяние волн на статистически неровной поверхности*, Наука, Москва, 1972.
16. Бенджамин Р., "Последние достижения в технике генерирования и обработки радиолокационных сигналов", *Зарубежная радиоэлектроника*, 1965, 19, **7**, 22-48.
16a. Берноскуни Ю. В., Вайсбург Г. М., "Система когерентного приема разнесенных сигналов Сатурн-К", *Труды НИИР*, 1976, **3**, 57.
17. Блейхман В. С., Брусиловский К. А., "Кодовые последовательности для испытания дискретных систем связи", *Электросвязь*, 1964, **12**, 56-63.
18. Блох Э. Л., Харкевич А. А., "Кодирование устойчивое по отношению к замиранию", *Электросвязь*, 1960, **4**, 3-6.
19. Блох Э. Л. и др., *Модели источника ошибок в каналах передачи цифровой информации*, Связь, Москва, 1971.
20. Блох Э. Л., Зяблов В. В., "Каскадные итерированные коды и применение их для исправления пакетов ошибок", в кн.: *Передача дискретных сообщений по каналам с группирующимися ошибками*, Наука, Москва, 1972.
21. Блох Э. Л., Зяблов В. В., "О существовании линейных двоичных кас-

кадных кодов с оптимальными корректирующими свойствами", *Проблемы передачи информации*, 1973, IX, 4, 3-10.

22. Блох Э. Л., Зяблов В. В., "Потенциальные и реализуемые корректирующие свойства каскадных кодов на основе кодов Рида-Соломона", в кн.: *Повышение верности передачи цифровой информации по дискретным каналам*, Наука, Москва, 1974.
23. Блох Э. Л., Зяблов В. В., "Кодирование обобщенных каскадных кодов", *Проблемы передачи информации*, 1974, X, 3, 45-50.
24. Блох Э. Л., Зяблов В. В., *Обобщенные каскадные коды*, Связь, Москва, 1976.
25. Блох Э. Л., Зяблов В. В., "Построение обобщенных каскадных кодов на базе кодов БЧХ", в кн.: *Кодирование и передача дискретных сообщений в системах связи*, Наука, Москва, 1976.
26. Блох Э. Л., Зяблов В. В., "Обобщенные каскадные коды", в кн.: *Актуальные проблемы теории информации*. Сер. Вопросы кибернетики, Научн. совет по компл. проблеме "Кибернетика", 1977, 29, 3-28.
27. Богданович В. А., "Способ построения подобных правил обнаружения сигналов при априорной неопределенности", *Радиотехника и электроника*, 1970, 7.
28. Большаков И. А., *Статистические проблемы выделения потока сигналов из шума*, Сов. радио, Москва, 1969.
29. Большаков И. А., Левин Б. Р., Репин В. Г., Тартаковский Г. П., "Некоторые вопросы статистического синтеза информационных систем", *Техническая кибернетика*, 1970, 2, 153-170.
30. Бомштейн Б. Д., Киселев Л. К., Моргачев К. Т., *Методы борьбы с помехами в каналах проводной связи*, Связь, Москва, 1975.
31. Борн М., *Моя жизнь и взгляды*, Прогресс, Москва, 1973.
Born, M., *My Life and My Views*, Scribner, N.Y., 1962.
32. Бородин Л. Ф. *Введение в теорию помехоустойчивого кодирования*, Сов. радио, Москва, 1968.
33. Бородич С. В., "О помехоустойчивости связи с импульсно-кодовой модуляцией", *Радиотехника*, 1949, 5, 13-27.
34. Бородич С. В., "Расчет шумов в каналах радиорелейных линий с частотным уплотнением и частотной модуляцией", *Электросвязь*, 1956, 1, 10-20; 3, 13-20.
35. Бородич С. В., *Искажения и помехи в многоканальных системах радиосвязи с частотной модуляцией*, Связь, Москва, 1976.
36. Бунимович В. И., "Преобразование флуктуаций нелинейной системой", *ЖТФ*, 1946, т. 16.
37. Бунимович В. И., Леонтович М. А., "О распределении числа больших отклонений при электрических флуктуациях", *ДАН АН СССР*, 1946, т. 53.
38. Бунимович В. И., *Флуктуационные процессы в радиоприемных устройствах*, Сов. радио, Москва, 1951.
39. Бухараев Р. Г., "Вопросы представимости в вероятностном автомате расширенных каналов связи и вероятностных последовательностей", *Итоговая конференция Казанского гос. ун-та за 1963*, Казань, 1964.
40. Быховский М. А., Рабинович В. С., "Помехоустойчивость многостанционной системы связи со сложными сигналами", *Труды НИИР*, 1967, 2, 82-90.
41. Быховский М. А., Рабинович В. С., Потенциальная помехоустойчивость многостанционных систем связи, использующих сигналы с большой базой и многопозиционное кодирование", *Труды НИИР*, 1968, 4, 14-24.
42. Быховский М. А., "Принципы построения устройств разнесенного приема ЧМ сигналов", *Электросвязь*, 1976, 4, 17-24.
43. Мурадян А. Г. (ред.), *Световоды с дискретной коррекцией для передачи информации*, Связь, Москва, 1975.

44. Вайнштейн Л. А., Зубаков В. Д., *Выделение сигналов на фоне случайных помех*, Сов. радио, Москва, 1960.
Wainstein, L. A. and Zubakov V. D., *Extraction of Signals from Noise*, Prentice Hall Inc., Englewood Cliffs, N.J., 1962.
45. Величкин А. И., *Теория дискретной передачи и непрерывных сообщений*, Сов. радио, Москва, 1970.
46. Вакман Д. Е., *Сложные сигналы и принцип неопределенности в радиолокации*, Сов. радио, Москва, 1965.
47. Вакман Д. Е., *Регулярный метод синтеза ФМ сигналов*, Сов. радио, Москва, 1967.
48. Вакман Д. Е., Седлецкий Р. М., *Вопросы синтеза радиолокационных сигналов*, Сов. радио, Москва, 1973.
49. Варакин Л. Е., Пышкин И. М., "О помехоустойчивости асинхронно-адресной системы связи со сложными сигналами и частотно-временным кодированием", *Труды учебных институтов связи*, 1967, **35**, 99-108.
50. Варакин Л. Е., Пышкин И. М., "К вопросу применения сложных сигналов в адресных системах связи", *Электросвязь*, 1967, **1**, 72-77.
51. Варакин Л. Е., *Теория сложных сигналов*, Сов. радио, Москва, 1970.
52. Варакин Л. Е., "Оптимальные фазоманипулированные сигналы", *Радиотехника*, 1971, **11**, 7-16.
53. Варакин Л. Е., "Выбор систем сигналов для ААСС при когерентном приеме", *Электросвязь*, 1971, **12**, 51-59.
54. Варакин Л. Е., Власов В. Н., "Системы дискретных частотно-модулированных сигналов", *Радиотехника и электроника*, 1972, **5**, 963-971.
55. Варакин Л. Е., "Сельская радиосвязь и кодовое разделение каналов", *Электросвязь*, 1973, **10**, 64-70.
56. Варакин Л. Е., "Свойства полного кода", *Радиотехника*, 1973, **8**, 1-6.
57. Варакин Л. Е., Волков Л. Н., "Корреляционные свойства сегментов *М*-последовательностей", *Радиотехника*, 1973, **2**, 23-28.
58. Варакин Л. Е., Власов В. Н., "Системы дискретных многочастотных сигналов", *Электросвязь*, 1974, **7**, 58-63.
59. Варакин Л. Е., Сперанский В. С., "К вопросу статистического выбора систем фазоманипулированных сигналов", *Труды учебных институтов связи*, 1974, **66**, 158-163.
60. Варакин Л. Е., Житков В. В., "Дисперсия полного произвольного кода", *Радиотехника*, 1976, **5**, 97-98.
61. Варакин Л. Е., Моисеева Г. Г., "Распределение периодических корреляционных функций фазоманипулированных сигналов", *Радиотехника*, 1976, **12**, 12-17.
62. Варакин Л. Е., Моисеева Г. Г., "Последовательности максимальной вероятности", *Труды учебных институтов связи*, 1976, **80**, 49-56.
63. Варакин Л. Е., "Совпадение структурных помех в радиотехнических системах с дискретными частотными сигналами", *Радиотехника и электроника*, 1976, **11**, 2422-2426.
64. Варакин Л. Е., Матвеева О. В., "Статистические свойства апериодических корреляционных функций дискретных частотных сигналов", *Известия вузов*. Сер. Радиоэлектроника, 1977, **3**, 46-49.
65. Варакин Л. Е., "Статистические свойства дискретных частотных сигналов с частотной манипуляцией", *Радиотехника*, 1977, **9**, 28-31.
66. Варакин Л. Е., "Выбор систем сигналов для ААСС при некогерентном приеме", *Труды учебных институтов связи*, 1977, **82**, 82-86.
67. Варакин Л. Е., Житков В. В., "Статистические свойства полного троичного кода", *Труды учебных институтов связи*, 1977, **82**, 75-81.
68. Варшамов Р. Р., "Оценка числа сигналов в кодах с коррекцией ошибок", *ДАН СССР*, 1957, **111**, 5.
69. Васильев Б. М., Молодцова Л. И., Николаев В. Ф., и др., "Система сбора и обработки данных на межпланетных станциях Марс-6 и Марс-7", *Труды*

6-го симпозиума по избыточности в информационных системах, ч. 1, ЛИАП, Ленинград, 1974.
70. Венедиктов М. Д., Марков В. В., Эйдус Г. Г., *Асинхронные адресные системы связи*, Связь, Москва, 1968.
71. Венедиктов М. Д., Женевский Ю. П., Марков В. П., Эйдус Г. Г., *Дельта-модуляция. Теория и применение*, Связь, Москва, 1976.
72. Венедиктов М. Д., Даниэлян С. А., Марков В. В., Эйдус Г. Г., *Многостанционный доступ в спутниковых системах связи*, Связь, Москва, 1973.
73. Возенкрафт Дж. М., "Последовательный прием при связи через канал с параметрами, изменяющимися во времени", в кн.: Багдади Е. Дж. (ред.), *Лекции по теории систем связи*,— русский перевод под ред. Б. Р. Левина, Мир, Москва, 1964.
Baghdady, E. J. (Ed.), *Lectures on Communication System Theory*, Mc. Graw-Hill, N.Y., 1961.
74. Воронин А. А., "К вопросу о потенциальной помехоустойчивости в каналах со случайным изменением параметров", *Электросвязь*, 1961, 10, 11-18.
75. Гельфанд И. М., "Обобщенные случайные процессы", *ДАН СССР*, 1955, 110, 5, 853-856.
76. Герасимов В. В., Охтяркин Е. Г., "Одновременное действие радиопомех и флуктуационного шума на приемник ЧМ сигналов", *Электросвязь*, 1969, 8, 10-14.
77. Герасимов В. В., Охтяркин Е. Г., Титков В. А., "О необходимой избирательности приемника ЧМ сигналов", *Электросвязь*, 1972, 9, 9-15.
78. Гиббс Дж. В., *Основные принципы статистической механики*, русский перевод под ред. К. В. Никольского, Гостехиздат, Москва, 1946.
Gibbs, J. W., *The Collected Works*, vol. 2, New York, 1936.
79. Гихман И. И., Скороход А. В., *Теория случайных процессов*, Наука, Москва, т. 1—1971, т. 2—1973.
80. Гинзбург С. А., Мурадян А. Г., "Линейные выравниватели в оптической кабельной линии", *Радиотехника и электроника*, 1978, 1, 48-73.
81. Глазов Б. И., "Спектры и корреляционные функции частотно-манипулированных шумоподобных сигналов", *Радиотехника*, 1970, 9, 4-10.
82. Глазов Б. И., "Числовые и периодические последовательности для формирования шумоподобных сигналов с частотной модуляцией", *Радиотехника*, 1972, 3, 91-93.
83. Глазов Б. И., Котенко Л. П., Мерзликин Б. С., "Анализ спектральных свойств L-ичных дискретных ЧМ сигналов", *Радиотехника*, 1975, 12, 15-19.
84. Гленн В., "Системы связи с кодовым уплотнением каналов", *Зарубежная радиоэлектроника*, 1965, 19, 3, 12-22.
85. Глобус И. А., "К вопросу о распределении полного двоичного кода", *Радиотехника*, 1977, 3, 80-82.
86. Глушков В. А., Калиниченко Л. А., Лазарев В. Г., Сифоров В. И., *Сети ЭВМ*, Связь, Москва, 1977.
87. Гольдштейн Ю. А., "Помехоустойчивость приема дискретной информации в канале с коэффициентом передачи, подчиняющимся m-распределению", *Электросвязь*, 1965, 10, 71-73.
88. Голубев В. Н., *Частотная избирательность радиоприемников АМ сигналов*, Связь, Москва, 1970.
89. Гоппа В. Д., "Новый класс линейных корректирующих кодов", *Проблемы передачи информации*, 1970, 3, 24-30.
90. Гуревич В. Э., Лопушнян Ю. Г., Рабинович Г. В., *Импульсно-кодовая модуляция в многоканальной телефонной связи*, А., Связь, Москва, 1973.
91. Гусятинский И. А., Рыскин Э. Я., "Теоретическое и экспериментальное исследование мощности переходных помех при многолучевом приеме", *Электросвязь*, 1962, 12, 3-13.
92. Гусятинский И. А., Рыскин Э. Я., "Теоретическое и экспериментальное

исследование флуктуаций амплитуды и фазы модулирующего сигнала в многолучевом канале с ЧМ", *Электросвязь*, 1965, 2, 24-33.

93. Гусятинский И. А., Немировский А. С., "Автокорреляционная система борьбы с замираниями сигнала на тропосферных линиях связи", *Электросвязь*, 1973, 2, 7-8.
94. Гутин В. С., "Корреляционные свойства случайных двоичных последовательностей", *Радиотехника и электроника*, 1973, 2, 409-412.
95. Гуткин Л. С., *Теория оптимальных методов радиоприема при флуктуационных помехах*, Госэнергоиздат, Москва, 1961.
96. Гуткин Л. С., *Современная радиоэлектроника и ее проблемы*, Сов. радио. Москва, 1968.
97. Введенский Б. А. (ред.), *Дальнее тропосферное распространение УКВ*, Сов. радио, Москва, 1965.
98. Денисов Н. Г., "О дифракции волн на хаотическом экране", *Известия вузов.* Сер. Радиофизика, 1961, IV, 4, 630-638.
99. Добрушин Р. Л., "Общая формулировка основных теорем Шеннона", *Успехи математических наук*, 1959, 14, 6, 3-104.
100. Добрушин Р. Л., "О последовательном декодировании методом Возенкрафта-Рейффена", в кн.: *Проблемы кибернетики*, Наука, Москва, 1964.
101. Дьячков В. И., Кантор Л. Я., "Об оптимальном проектировании систем спутниковой связи с частотным многостанционным доступом", *Электросвязь*, 1969, 3, 27-32.
102. Максимов М. В. (ред.), *Защита от радиопомех*, Сов. радио, Москва, 1976.
103. Зигангиров К. Ш., "Некоторые последовательные процедуры декодирования", *Проблемы передачи информации*, 1966, II, 4, 13-25.
104. Зигангиров К. Ш., *Процедуры последовательного декодирования*, Связь, Москва, 1974.
105. Зигангиров К. Ш., "Алгоритм последовательного декодирования с возвращениями", в кн.: *Повышение верности передачи цифровой информации по дискретным каналам*, Связь, Москва, 1974, 43-48.
106. Зиновьев В. А., "Обобщенные каскадные коды", *Проблемы передачи информации*, 1976, XII, 1, 5-15.
107. Зиновьев В. А., Зяблов В. В., "Декодирование нелинейных обобщенных каскадных кодов", Четвертый международный симпозиум по теории информации. Тезисы докладов, ч. II, М.-Л., 1976, 42-44.
108. Зиновьев В. А., Зяблов В. В., "Исправление обобщенными каскадными кодами независимых ошибок при наличии пакетов ошибок", Седьмой всесоюзный симпозиум по проблеме избыточности в информационных системах. Тезисы докладов, ч. I, ЛИАП, Ленинград, 1977.
109. Зюко А. Г., *Помехоустойчивость и эффективность систем связи*, Связь, Москва, 1972.
110. Зяблов В. В., "Анализ корректирующих свойств итерированных и каскадных кодов", в кн.: *Передача цифровой информации по каналам с памятью*, Наука, Москва, 1970.
111. Зяблов В. В., "Оценка сложности построения двоичных линейных каскадных кодов", *Проблемы передачи информации*, 1971, VII, 1, 5-13.
112. Зяблов В. В., "Оптимизация алгоритмов каскадного декодирования", *Проблемы передачи информации*, 1973, IX, 1, 26-32.
113. *Инженерно-технический справочник по электросвязи*, "Радиорелейные линии", Связь, Москва, 1971.
114. Каган Б. Д., Финк Л. М., "Метод последовательного приема в целом для кодов, допускающих мажоритарное декодирование", *Электросвязь*, 1967, 1, 14-22.
115. Каваков И. Е., *Статистическая теория систем управления в пространстве состояний*, Наука, Москва, 1975.
116. Калашников Н. И., *Основы расчета электромагнитной совместимости систем связи через ИСЗ*, Связь, Москва, 1970.

117. Калашников Н. И., Мордухович Л. Г., "Энергетический спектр высокочастотного колебания, модулированного по частоте многоканальным телефонным сообщением", *Радиотехника*, 1973, 28, 8, 18-24.
118. Калашников Н. И., Степанов А. П., "К расчету радиопомех с амплитудной и угловой модуляцией в многоканальных системах связи с ЧМ", *Радиотехника*, 1974, 29, 2, 1-7.
119. Калашников Н. И., Меркадер Л. П., Тимищенко М. Г., Юдин А. И., *Системы связи и РРЛ*, Связь, Москва, 1977.
120. Калинин А. И., Черенкова Е. Л., *Распространение радиоволн и работа радиолиний*, Связь, Москва, 1971.
121. Каневский З. М., *Передача сообщений с информационной обратной связью*, Связь, Москва, 1969.
122. Кантор Л. Я., Дьячкова М. Н., Дорофеев В. М., "Влияние радиопомех на приемник ЧМ сигналов", *Электросвязь*, 1971, 6, 39-45.
123. Кантор Л. Я., "Новый подход к оценке эффективности использования геостационарной орбиты", *Электросвязь*, 1976, 1, 5-11.
124. Кантор Л. Я., Дорофеев В. М., *Помехоустойчивость приема ЧМ сигналов*, Связь, Москва, 1977.
125. Катюк А. Ф., Ольшевский В. В., Цветков Э. И., *Методы и аппаратура для анализа характеристик случайных процессов*, Энергия, Москва, 1967.
126. Шварцман В. О. (ред.), *Каналы передачи данных*. Связь, Москва, 1970.
127. Кеннеди Р., "Введение в теорию передачи сообщений по оптическим каналам с рассеянием", *ТИИЭР*, 1970, 58, 10, 264-278.
Kennedy, R. S., "Communication Through Optical Scattering Channels: An Introduction", *IEEE*, 1970, 58, 10, 1651-1665.
128. Кириллов И. Е., *Помехоустойчивая передача сообщений по линейным каналам со случайно меняющимися параметрами*, Связь, Москва, 1971.
129. Кириллов И. Е., Сойфер В. А., "Пространственно-временные характеристики линейных каналов с переменными параметрами", *Проблемы передачи информации*, 1972, VIII, 2, 40-46.
130. Кловский Д. Д., "Вопросы потенциальной помехоустойчивости при замираниях сигнала", *Радиотехника*, 1960, 5, 17-25.
131. Кловский Д. Д., "О потенциальной помехоустойчивости коротковолновой радиотелеграфии", *Электросвязь*, 1960, 9, 3—11.
132. Кловский Д. Д., *Передача дискретных сообщений по радиоканалам*, Связь, Москва, 1969.
133. Кловский Д. Д., *Теория передачи сигналов*. Связь, Москва, 1973.
134. Кловский Д. Д., Николаев Б. И., *Инженерная реализация радиотехнических схем (в системах передачи дискретных сообщений в условиях межсимвольной интерференции)*, Связь, Москва, 1975.
Klovsky. D. and B. Nikolaev, *Sequential Transmission of Digital Information in the Presence of Intersymbol Interference*, Mir, Moscow, 1978.
135. Кловский Д. Д., Сойфер В. А., *Обработка пространственно-временных сигналов*, Связь, Москва, 1976.
136. Клюев Н. И., *Информационные основы передачи сообщений*, Сов. радио, Москва, 1966.
137. Козин И. В., *Элементы теории оптимального обнаружения и приема сигналов*, изд. Ленинградского ун-та, Ленинград, 1974.
138. Колесник В. Д., Мирончиков Е. Г., *Декодирование циклических кодов*, Связь, Москва, 1968.
139. Колмогоров А. Н., *Основные понятия теории вероятностей*, Наука, Москва, 1974.
Kolmogorov, A. N., *Foundations of the Theory of Probability*, Chelsea Publishing Co., New York, N.Y., 1950.
140. Колмогоров А. Н., "Интерполирование и экстраполирование стационарных случайных последовательностей", *Известия АН СССР*, серия математическая, 1941, 5, 5.

141. Колмогоров А. Н., "Теория передачи информации", сборник: Сессия АН СССР по научным проблемам автоматизации производства, 1956, *изд. АН СССР*, Москва, 1957.
142. Коновалов Г. В., Тарасенко Е. М., *Импульсные случайные процессы в электросвязи*, Связь, Москва, 1973.
143. Кор, Кручфилд, Мерчиз, "Импульсная УКВ станция, использующая шумоподобные сигналы", *Зарубежная радиоэлектроника*, 1966, 20, 4, 20-31.
144. Корадо В. А., "Об оптимальном обнаружении сигналов на фоне помех", *Радиотехника и электроника*, 1972, 1, 173-175.
145. Корбут А. А., Финкельштейн Ю. Ю., *Дискретное программирование*, Наука, Москва, 1969.
146. Коржик В. И., Осмоловский С. А., Финк Л. М., "Универсальное кодирование для произвольных каналов с обратной связью", *Проблемы передачи информации*, 1974, 4, 25-29.
147. Коржик В. И., Финк Л. М., *Помехоустойчивое кодирование дискретных сообщений в каналах со случайной структурой*, Связь, Москва, 1975.
148. Коржик В. И., Финк Л. М., "Многоступенчатое стохастическое кодирование", *Проблемы передачи информации*, 1978, 2.
149. Коноплева Е. Н., "Кривые распределения напряженности поля коротковолновых сигналов", *Электросвязь*, 1959, 9, 20-27.
150. Костас Дж., "Пропускная способность каналов с замираниями в условиях сильных помех", пер. с англ.: *ТИИЭР*, 1963, 51, 3, 451-460.
Costas, J. P., "Information Capacity of Fading Channels Under Conditions of Intense Interference", *IEEE*, 1963, 51, 3, 451-461.
151. Котельников В. А., *Теория потенциальной помехоустойчивости*, ГЭИ, Москва, 1956.
152. Красовский А. А., *Фазовое пространство и статистическая теория динамических систем*, Наука, Москва, 1974.
153. Крейн М. Г., "Об основной аппроксимационной задаче теории экстраполяции и фильтрации стационарных случайных процессов", *ДАН*, 1954, 94, 13-16.
154. Кремер И. Я., Владимиров В. И., Карпухин В. И., *Моделирование помехи и прием радиосигналов*, Сов. радио, Москва, 1972.
155. Кузнецов А. В., Цыбаков Б. С., "Кодирование в памяти с дефектными ячейками", *Проблемы передачи информации*, 1974, 2.
156. Кузнецов В. П., "Инвариантность решений по отношению к мешающему параметру", *Проблемы передачи информации*, 1971, 4, 36-44.
157. Кумыш Э. И., "Оценка помехозащищенности передачи цветных и черно-белых телевизионных изображений в области порога выигрыша ЧМ", в кн.: *Методы помехоустойчивости приема ЧМ и ФМ сигналов*, Сов. радио, Москва, 1976.
158. Кушнир А. Ф., Левин Б. Р., "Оптимальные ранговые алгоритмы обнаружения сигналов в шумах", *Проблемы передачи информации*, 1968, 3, 3-18.
159. Кушнир А. Ф., Левин Б. Р., "Ранговые алгоритмы обнаружения сигналов по зависимой выборке", *Радиотехника*, 1971, 4, 38-42.
160. Кушнир А. Ф., Пинский А. И., "Асимптотически оптимальные критерии для проверки гипотез при зависимой выборке наблюдений", *Теория вероятности и ее применения*, 1971, 16, 2, 280-290.
161. Левин Б. Р., *Теория случайных процессов и ее применение в радиотехнике*, Сов. радио, Москва, 1957.
162. Левин Б. Р., *Теоретические основы статистической радиотехники*, кн. первая, Сов. радио, Москва, 1-е изд.—1966, 2-е изд.—1974.
163. Левин Б. Р., *Теоретические основы статистической радиотехники*, кн. вторая, Сов. радио, Москва, 1-е изд.—1968, 2-е изд.—1975.

164. Левин Б. Р., *Теоретические основы статистической радиотехники*, кн. третья, Сов. радио, Москва, 1976.
165. Левин Б. Р., Левин Г. А., Айзенберг В. И., Розанов В. С., "Повышение эффективности многоканальных систем с временным разделением каналов", *Электросвязь*, 1960, 5, 10-16.
166. Левин Б. Р., Розанов В. С., "Расчет числа каналов многоканальных систем с интервальной время-импульсной модуляцией", *Электросвязь*, 1961, 6, 10-14.
167. Левин Б. Р., "Актуальные проблемы статистического синтеза информационных систем", *Радиотехника*, 1971, 4, 5-7.
168. Левин Б. Р., Кушнир А. Ф., "Асимптотически оптимальные алгоритмы обнаружения и различения сигналов на фоне помех", *Радиотехника и электроника*, 1969, 2, 249-258.
169. Левин Б. Р., Кушнир А. Ф., "Асимптотически оптимальные ранговые алгоритмы обнаружения сигналов на фоне помех", *Радиотехника и электроника*, 1969, 2, 259-266.
170. Левин Б. Р., Рыбин А. К., "Непараметрические амплитудные и фазовые методы обнаружения сигналов ч. I", *Техническая кибернетика*, 1969, 5, 110-118.
171. Левин Б. Р., Рыбин А. К., "Непараметрические амплитудные и фазовые методы обнаружения сигналов, ч. II", *Техническая кибернетика*. 1970, 1, 153-160.
172. Левин Б. Р., "Оптимальные алгоритмы обнаружения сигналов, устойчивые к изменению априорных данных", *Известия вузов*. Сер. Радиоэлектроника, 1970, XIII, 2, 109-121.
173. Левин Б. Р., Кушнир А. Ф., Пинский А. И., "Асимптотически оптимальные алгоритмы обнаружения и различения сигналов на фоне коррелированных помех", *Радиотехника и электроника*, 1971, 5, 743-754.
174. Левин Б. Р., Баронкин В. М., "Обнаружение сигналов на фоне помех по квантованным наблюдениям", *Радиотехника и электроника*, 1973, 5, 940-949.
175. Левин Б. Р., Баронкин В. М., "Синтез асимптотически оптимальных алгоритмов обнаружения на фоне помех", *Радиотехника и электроника*, 1974, 5, 1017-1040.
176. Левин Б. Р., "Асимптотически оптимальные алгоритмы обнаружения сигналов на фоне помех (обзор)", *Труды СФТИ*, 1973, 63.
177. Левин Б. Р., Шварц В., "Об определении распределения процесса на выходе линейной системы", *Труды учебных институтов связи*, 1976. 81.
178. Левин Б. Р., Шинаков Ю. С., "Совместно оптимальные алгоритмы обнаружения сигналов и оценивания их параметров (обзор)"; *Радиотехника и электроника*, 1977, 11, 2256-2259.
179. Левин Б. Р. (ред.), *Лекции по теории систем связи*, пер. с англ., Мир, Москва, 1969.
Baghdady, E. J. (Ed.), *Lectures on Communication System Theory*, McGraw-Hill, New York, N.Y., 1961.
180. Лившиц А. Р., Биленко А. П., *Многоканальные асинхронные системы передачи информации (элементы теории)*, Связь, Москва, 1974.
181. Липцер Р. Ш., Ширяев А. Н., *Статистика случайных процессов. Нелинейная фильтрация и смежные вопросы*. Наука, Москва, 1974.
182. Лобкова Л. М., Литвинова Т. П., Милютин Е. Р. и др., "Экспериментальное исследование законов распределения флуктуаций мощности лазерного излучения в турбулентной атмосфере", *Труды учебных институтов связи*, 1969, 46, 54-61.
183. Лубны-Герцык А. Л., "Расчет кросс-модуляции при многоканальной передаче с очень большим числом каналов", *Электросвязь*, 1939, 6.
184. Ляшенко И. Н., Карагадова Е. А., Черникова М. В., Шор Н. З., *Линейное и нелинейное программирование*, Вища школа, Киев, 1975.

185. Марченко Ю. Ф., "К расчету допустимой величины радиопомех с амплитудной модуляцией в радиорелейных системах", *Электросвязь*, 1968, 3, 11-18.
186. Мельников В. С., *Частотное радиотелеграфирование*, Связьиздат, Москва, 1952.
187. Меньшиков Г. Г., *Двоичная аппроксимация: основы теории, применение и вопросы передачи сообщений*, ЛЭИС, Ленинград, 1968.
188. Мешковский К. А., "Оптимальные и близкие к ним двоичные коды", *Электросвязь*, 1958, 5, 5-15.
189. Мешковский К. А., Кириллов Н. Е., *Кодирование в технике связи*, Связь, Москва, 1966.
190. Мирский Г. Я., *Аппаратурное определение характеристик случайных процессов*, 2-е изд., Энергия, Москва, 1972.
191. Мирский Г. Я., *Радиоэлектронные измерения*, Госэнергоиздат, Москва, 1963.
Mirsky, G., *Radioelectronic Measurements*, Mir, Moscow, 1978.
192. *Материалы XIII Пленарной Ассамблеи МККР*, т. IV и IX (на англ., фран. и исп. яз.), изд. Международного Союза Электросвязи, Женева, 1974.
193. Моисеева Г. Г., "Построение больших производных систем ФМ сигналов", *Электросвязь*, 1977, 6, 67-72.
194. Морроу, "Общая характеристика каналов", в кн.: *Лекции по теории систем связи*, русский перевод под ред. Б. Р. Левина, Мир, Москва, 1964.
195. Мурадян А. Г., Гинзбург С. А., "Параметры высокоскоростных систем передачи по оптическим кабелям", *Электросвязь*, 1976, 5, 56-60.
196. Мурадян А. Г., Гинзбург С. А., "Оптимальный фильтр для приемника на световоде", *Квантовая электроника*, 1977, 4, 5, 1147-1149.
197. Немировский А. С., "Методы приема и комбинирования разнесенных сигналов", *Сб. трудов НИИ Мин. связи*, 1960, 2 (20).
198. Немировский А. С., Тараканова Т. С., "Распределение отношения сигнал/шум в телефонном канале тропосферной линии при разнесенном приеме с учетом порога ЧМ", *Электросвязь*, 1970, 4, 78-79.
199. Нортон, Воглер, Мансфилд, Шорт, "Вероятностное распределение суммарной амплитуды постоянного вектора и вектора, распределенного по закону Рэлея", в кн.: *Вопросы дальней связи на УКВ*, Сов. радио, Москва, 1957.
200. Носов Ю. Р., Шилин В. А., *Полупроводниковые приборы с зарядовой связью*, Сов. радио, Москва, 1976.
201. Овчинников В. В., Бреусов В. И., "О последовательном декодировании в каналах с памятью", Материалы V конференции по теории передачи и кодирования информации, секция 6, Научный совет по комплексной проблеме "Кибернетика", Москва-Горький, 1972.
202. Окунев Ю. Б. "Системы связи, инвариантные к помехам", *Радиотехника*, 1971, 8, 1-7.
203. Ольховский Ю. Б., Новоселов О. Н., Мановцев А. П., *Сжатие данных при телеизмерениях*, Сов. радио, Москва, 1971.
204. Остроухов В. С., Тузов Г. И., "Исследование функции неопределенности сигнала с частотно-фазовой манипуляцией", *Радиотехника и электроника*, 1974, 11, 2309-2313.
205. Остроухов В. С., Тузов Г. И., "Функция автокорреляции сигнала с частотно-фазовой манипуляцией", *Радиотехника и электроника*, 1974, 11, 2314-2320.
206. Остроухов В. С., Тузов Г. И., "Некоторые особенности нейтральной области автокорреляционной функции сигнала с частотной фазовой манипуляцией", *Радиотехника*, 1976, 2, 85-87.

207. Пашаев И. С., "Некоторые применения алгоритма эвристического ветвления", *ВИНИТИ*, 7758-73, 1973.
208. Петров Б. Н. и др., *Информационно-семантические проблемы в процессах управления и организации*, Наука, Москва, 1977.
209. Петрович Н. Т., "Новые способы осуществления фазовой телеграфии", *Радиотехника*, 1957, 10, 47-54.
210. Петрович Н. Т., Размахин М. К., *Системы связи с шумоподобными сигналами*, Сов. радио, Москва, 1969.
211. Пилипчук Н. И., Яковлев В. П., "Алгоритмы адаптивной дискретизации и их классификация", *Приборы управления*, 1977, 2, 3-5.
211а. Пинскер М. С., *Информация и информационная устойчивость случайных величин и процессов*, изд. АН СССР, 1960.
212. Полищук Ю. М., "О неравномерности статистического распределения фазы некогерентной составляющей рассеянных полей", *Радиотехника и электроника*, 1975, 5.
213. Попов О. В., "Исправление группирующихся ошибок методом каскадной локализации", в кн.: *Повышение верности передачи цифровой информации по дискретным каналам*, Наука, Москва, 1974, 177-183.
214. Поспелов Д. А., *Вероятностные автоматы*, Энергия, Москва, 1970.
215. Просин А. В., "К теории каналов радиосвязи со статистически неровными поверхностями", *Труды четвертого коллоквиума по УКВ связи*, ч. I, Будапешт, 1970.
216. Пугачев В. С., *Теория случайных функций и ее применение к задачам автоматического управления*, Гостехиздат, Москва, 1957.
Pugachev, V. S., *Theory of Random Functions and Its Application to Control Problems*, Pergamon Press Ltd., Oxford, England.
217. Пугачев В. С., "Оптимальные алгоритмы обучения автоматических систем", *ДАН АН СССР*, 1967, 172, 5, 1043-1045.
218. Пугачев В. С., Казаков И. Е., Евланов Л. Г., *Основы статистической теории автоматических систем*, Машиностроение, Москва, 1974.
219. Пуртов Л. П., Замрий А. С., Захаров А. И., "Основные закономерности распределения ошибок в дискретных каналах связи", *Электросвязь*, 1967, 2, 1-8.
220. Пышкин И. М., "Эффективность асинхронных адресных систем связи с кодовым разделением при передаче дискретной информации", *Электросвязь*, 1971, 25, 4, 28-30.
221. Пышкин И. М., Шауро А. В., "Эффективность асинхронных адресных систем связи при передаче телефонных сообщений", *Электросвязь*, 1972, 26, 10, 28-30.
222. Пышкин И. М., Чвилев Г. Д., "Помехоустойчивость и эффективность асинхронных адресных систем с частотно-временным кодированием при когерентном приеме дискретной информации", *Радиотехника и электроника*, 1973, 11, 2427-2430.
223. Радченко А. Н., "Кодовые кольца как способ представления кодовых множеств", *Автоматика и телемеханика*, 1959, 20, 7, 970-977.
224. Размахнин М. К., "Широкополосные системы связи", *Зарубежная радиоэлектроника*, 1965, 8, 3-29.
225. *Регламент радиосвязи*, Связь, Москва, 1975.
226. Репин В. Г., Тартаковский Г. П., *Статистический синтез при априорной неопределенности и адаптации информационных систем*, Сов. радио, Москва, 1977.
227. Рид, Блэсбалг, "Эффективные методы измерения расстояний и передачи данных в условиях многолучевого распространения на линиях самолет-земля и земля-самолет", *ТИИЭР*, 1970, 58, 3, 146-154.
Reed, I. S., Blasbalg, "Multipath Tolerant Ranging and Data Transfer Techniques for Air-to-Ground and Ground-to-Air Links", *IEEE*, 1970, 58, 3, 422-429.

228. Роббинс Г., "Эмпирический байесовский подход к задачам теории статистических решений", в кн.: *Математика*; пер. с англ., Мир. Москва, 1965, 5.
229. Романов В. Д. "Цифровые системы многоканальной связи и пути их развития. Итоги науки и техники", *Электросвязь*, 1977, 8, 5-51.
230. Рухин А. Л., Самсоненко С. В., "О процедуре обнаружения инвариантной относительно интенсивности сигнала и помехи", *Радиотехника и электроника*, 1972, 1, 170-172.
231. Рытов С. М., *Введение в статистическую радиофизику*, Наука, Москва, 1966.
232. Сагателов В. С., "Спектральные и корреляционные свойства составных двоичных последовательностей", *Радиотехника и электроника*, 1963, 1, 201-205.
233. Самойленко С. И., "Алгоритмы целенаправленного стохастического поиска", Доклад на III международной конференции по искусственному интеллекту. Стэнфорд США, *ВИНИТИ*, 111-74, Деп. 1973.
234. Самойленко С. И., *Системы обработки информации*, Наука, Москва, 1975.
235. Самойленко С. И., "Размытые эвристики", Доклад на 4-ой Международной конференции по искусственному интеллекту, Тбилиси, *ВИНИТИ*, 1877, Деп. 1975Б.
236. Самойленко С. И., "Субоптимальное программирование", в кн.: *Семиотика и информатика*, *ВИНИТИ*, Москва, 1977, 8, 3-44.
237. Сандерс Д., "Система связи Диджилок", в кн.: *Передача цифровой информации*, перев. с англ. под ред. С. И. Самойленко, ИЛ, Москва, 1963.
238. Свердлик М. Б. *Оптимальные дискретные каналы*, Сов. радио, Москва, 1975.
239. Свириденко С. С., *Основы синхронизации при приеме дискретных сигналов*, Связь, Москва, 1974.
240. Свириденко В. А., *Анализ систем со сжатием данных*, Связь, Москва, 1977.
241. Севальнев Л. А., Шендерович А. М., *Передача сигналов цветного телевизионного изображения по линиям связи*, Связь, Москва, 1973.
242. Сикарев А. А., "Оптимальный некогерентный прием в каналах с флуктуационными и сосредоточенными помехами", *Проблемы передачи информации*, 1970, 2, 109-118.
243. Сикарев А. А., Цыганков В. В., "Оптимальный прием при флуктуационных и сосредоточенных помехах адаптивными многополюсными фильтрами", *Радиотехника*, 1976, 10, 52-61.
244. Сифоров В. И., "О влиянии помех на прием импульсных сигналов", *Радиотехника*, 1946, 1, 5-19.
245. Сифоров В. И., "Исследование одновременного замирания двух радиосигналов на коротких волнах", Труды международной конференции по ВЧ радиовещанию в Мексике, 1949. Доклад от СССР в комитете по планированию. Мехико-сити, протокол 11А, доклад 79-Р
246. Скороход А. В., "Конструктивные методы задания случайных процессов", *Успехи математических наук*, 1965, XX, 3(123), 67-87.
247. Смирнов В. А., "Теоретическое изучение влияния многолучевого распространения радиоволн на КВ связь при ЧМ", *Журнал технической физики*, 1945, XV, 11, 815-832.
248. Смирнов Н. И., Голубков Н. А., "О свойствах составных последовательностей", *Радиотехника и электроника*, 1973, 1, 197-200.
249. Смирнов Н. И., "Применение *М*-последовательностей в асинхронных радиотехнических системах", *Электросвязь*, 1970, 10, 33-42.
250. Стратонович Р. Л. "К теории нелинейной фильтрации случайных функций", *Теория вероятностей и ее приложения*, 1959, 4, 239-247.
251. Стратонович Р. Л., "Применение теории марковских процессов для

оптимальной фильтрации сигналов", *Радиотехника и электроника*, 1960, 11, 1751-1763.
252. Стратонович Р. Л., "Условные процессы Маркова", *Теория вероятности и ее применения*, 1960, 5, 2, 172-195.
253. Стратонович Р. Л., *Избранные вопросы теории флуктуации в радиотехнике*, Сов. радио, Москва, 1961.
Stratonovich, R., *Topics in the Theory of Random Noise*, Gordon and Breach, New York, N.Y., 1963.
254. Стратонович Р. Л., *Условные марковские процессы и их применения к теории оптимального управления*, Изд. МГУ, Москва, 1966.
255. Стратонович Р. Л., *Принципы адаптивного приема*, Сов. радио, Москва, 1973.
256. Талызин Н. В., Кантор Л. Я., Манякин Е. А., Паянский Ю. М., "Об оптимальных параметрах и экономической эффективности многостанционной системы спутниковой связи", *Радиотехника*, 1969, 11, 5-13.
257. Тарасенко Ф. П., *Непараметрическая статистика*, Изд-во Томского университета, Томск, 1976.
258. Татарский В. И., *Распространение волн в турбулентной атмосфере*, Наука, Москва, 1968.
259. Титсворт, "Применение булевой функции для построения многоканальной телеметрической системы", *Зарубежная радиоэлектроника*, 1964, 18, 8, 33-39.
260. Титов В. П., "Взаимокорреляционные свойства бинарных последовательностей", *Радиотехника*, 1972, 4, 83-85.
261. Тихонов В. И., Кульман Н. К., *Нелинейная фильтрация и квазикогерентный прием сигналов*, Сов. радио, Москва, 1975.
261a. Тихонов В. И., "Влияние шумов на работу схемы фазовой автоподстройки частоты", *Автоматика и телемеханика*, 1959, 9, 1189-1196.
262. Тихонов В. И., Миронов М. А., *Марковские процессы*, Сов. радио, Москва, 1977.
263. Фельдбаум А. А., Дудыкин А. Д., Мановцев А. Б., Миролюбов Н. Н., *Теоретические основы связи и управления*, Физматгиз, 1964.
264. Уздемир А. П., "Корреляционные функции комбинированных последовательностей", *Радиотехника и электроника*, 1972, 3, 499-510.
265. Фалько А. И., "К вопросу подавления сосредоточенных помех в широкополосных системах связи", *Электросвязь*, 1969, 7, 9-14.
266. Фалько А. И., "Об оптимальном приеме при воздействии 'небелого' шума", *Радиотехника*, 1970, 8.
267. Финк Л. М., *Теория передачи дискретных сообщений*, Сов. радио, Москва, 1963.
268. Фельдбаум А. А., *Основы теории оптимальных автоматических систем*, Физматгиз, Москва, 1963.
269. Фортушенко А. Д., Быков В. Л., и др., *Основы технического проектирования систем связи через ИСЗ*, Связь, Москва, 1970.
270. Фортушенко А. Д., Афанасьев Ю. А., и др., *Основы технического проектирования аппаратуры систем связи с помощью ИСЗ*, Связь, Москва, 1972.
271. Халин Ф. М., Леонов А. Ф., Маладзе В. В., *Методы повышения качества электронных систем коммутации*, Связь, Москва, 1971.
272. Харкевич А. А., *Очерки общей теории связи*, Гостехиздат, Москва, 1955.
273. Харкевич А. А., *Борьба с помехами*, Физматгиз, Москва, 1964.
274. Харкевич А. А., *Избранные труды*, т. 3, Наука, Москва, 1973.
275. Хаффмен Д. А., "Исследование сигналов, эквивалентных импульсу", *Радиотехника*, 1964, 8, 3-8.
276. Хворостенко Н. П., *Статистическая теория демодуляции дискретных сигналов*, Связь, Москва, 1968.

277. Хинчин А. Я., "Об основных теоремах теории информации", *Успехи математических наук*, 1956, 11, 1.
278. Хинчин А. Я., *Работы по математической теории массового обслуживания*, Физматгиз, Москва, 1963.
Khinchine, Y., *Mathematical Methods in the Theory of Queueing*, Griffin, 1960.
279. Холланд, Клейборн, "Устройства на поверхностных акустических волнах", *ТИИЭР*, 1974, 5, 45-82.
Holland, M. G., Claiborne, L.T., "Practical Surface Acoustic Wave Devices", *IEEE*, 1974, 5, 582-611.
280. Хургин Я. И., Яковлев В. П., *Финитные функции в физике и технике*, Наука, Москва, 1972.
281. Цветнов В. В., "Статистические свойства сигналов и помех в двухканальных фазовых системах", *Радиотехника*, 1957, 5, 12-29.
282. Цыбаков Б. С., "Исправление дефектов и ошибок", *Проблемы передачи информации*, 1975, 3.
283. Цыпкин Я. З., *Основы теории обучающих систем*, Наука, Москва, 1970.
284. Чабдаров Ш. М., Трофимов А. Т., "Полигауссовы представления произвольных помех и прием дискретных сигналов", *Радиотехника и электроника*, 1975, 4, 734-745.
285. Чабдаров Ш. М., "Оптимальный прием дискретных сигналов при комплексе шумовых и импульсных помех", *Радиотехника и электроника*, 1977, 6, 1162-1174.
286. Чабдаров Ш. М., "Полигауссовы приемники произвольно флуктуирующих сигналов", *Известия вузов СССР*, сер. Радиоэлектроника, 1977, XX, 9, 32-38.
287. Чернов Л. А., *Распространение волн в среде со случайными неоднородностями*, изд. АН СССР, Москва, 1958.
288. Черняк Ю. Б., "Чувствительность, точность и разрешающая способность многоканального приемника с широкополосным ограничителем", *Радиотехника и электроника*, 1962, 8, 1302-1310.
289. Черняков М. В., Ярлыков М. С., "Устройство оптимальной нелинейной фильтрации в приемниках асинхронно-адресных систем связи для сигналов с время-импульсной модуляцией", *Радиотехника*, 1974, 29, 9, 6-12.
290. Черняков М. В., "Синтез оптимальных приемников асинхронно-адресной системы связи с внутриимпульсной фазовой манипуляцией", *Радиотехника*, 1976, 2, 11-18.
291. Черняков М. В., "Оптимальная фильтрация сигналов в асинхронных системах связи между подвижными объектами", *Радиотехника*, 1977, 1, 21-28.
292. Черняков М. В., "Оптимальный прием сигналов с время-импульсной модуляцией в асинхронных адресных системах связи с частотно-временным кодированием на фоне внутрисистемных помех", *Радиотехника*, 1977, 6, 16-23.
293. Чиби Ш., "Непосредственная интерференция между расположенными близко друг к другу радиорелейными каналами с частотным уплотнением и частотной модуляцией", *Acta Technica Acad. Sc. Hung.*, 1963, vol. 42, 7-18.
294. Шапцев В. А., "Помехоустойчивость некогерентного приема в канале с гамма-замираниями", *Труды сибирского физико-технического института*, 1973, 64.
295. Шахгильдян В. В., Ляховкин А. А., *Системы фазовой автоподстройки частоты*, Связь, Москва, 1972.
296. Шахгильдян В. В., Лохвицкий М. С., *Методы адаптивного приема сигналов*, Связь, Москва, 1974.
297. Шерман Г., "Быстрые мультипликативные флуктуации", в кн.: *Лекции*

по теории систем связи, русский перевод под ред. Б. Р. Левина, Мир, Москва, 1964.
Baghdady, E. J. (Ed.), *Lectures on Communication System Theory*, McGraw-Hill, New York, N.Y., 1961.
298. Ширман Я. Д., *Разрешение и сжатие сигналов*, Сов. радио, Москва, 1974.
299. Шляхов И. М., "Вопросы адаптивного приема сигналов в канале с флуктуационным шумом и сосредоточенными помехами", *Радиотехника*, 1976, 8, 10-18.
300. Щукин А. Н., "Об одном методе борьбы с импульсными помехами", Известия АН СССР. Сер. физическая, 1946, 10, 1.
301. Ярлыков М. С., Черняков М. В., "Субоптимальный прием некогерентных сигналов в асинхронно-адресных системах с частотно-временным кодированием", *Проблемы передачи информации*, 1973, IX, 2, 43-52.
302. Ярлыков М. С., Черняков М. В., "Квазиоптимальный прием сигналов в асинхронных адресных системах связи с частотно-временной матрицей", *Известия вузов СССР*. Сер. Радиоэлектроника, 1973, XVI, 5, 42-50.
303. Ярлыков М. С., Черняков М. В., "Оптимальный прием сигналов с время-импульсной модуляцией в асинхронно-адресных системах связи", *Проблемы передачи информации*, 1974, X, 4, 56-64.
304. Ярлыков М. С., Черняков М. В., "Оптимальная обработка сигналов в асинхронных адресных системах передачи информации с учетом подвижности абонентов", Проблемы управления и теории информации, 1977, 6, 6.
305. Abramson N., "The ALOHA system—another alternative for computer communication", *AFIPS Conf. Proc.*, 1970, vol. 37.
306. Abramson N., "The ALOHA system", in *Computer communication Network*, Prentice-Hall, N.Y., 1973.
307. Abramson N., Kuo F., *Computer-communication networks*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
308. Abramson N., "The theory of packet broadcasting. The ALOHA system, B76-1", January 1976, Univ. of Hawaii.
309. Abramson N., "A class of systematic codes for non-independent errors", IRE Trans. 1959, IT-5, 150-157.
310. *Advances in communication systems*, A.V. Balakrishnan (Ed.) Acad. Press. N.-Y., 1965.
311. Agnew C., "On quadrate adaptive routing alogrithms". *Commun. Ass. Comput. Mach.* 1976, vol. 19, 18-22.
312. Algazi V. R., "Useful approximation to opimum quantization". *IEEE Trans.*, 1966, COM-14, 1, 297-301.
313. Allen W. B., Westerfield E. C., "Digital compressed-time correlators and matched filters for active sonar," *J. Am. Acoust. Soc.* 1964. 36.
314. Assadowrian F., Jacoby D. L., "Multiple access consideration for communication satellites," RCA Rev. 1966, 27.
315. Balakrischnan A. V., *Communication theory*. McGraw-Hill. N.Y., 1968.
316. Barucha-Reid A. T., *Elements of the theory of stochastic processes and their applications*, McGraw-Hill, N.Y., 1960.
317. Barker R. H., "Group synchronization of lineary digital systems," in *Communication Theory*, Academic Press. N.Y., London, 1953.
318. Battiu R. N., "A statistical optimizing navigation procedure for space flight," *J. Amer. Rocket Soc.*, 1962, vol. 32.
319. Beckman P., "Statistical distribution of the amplitude and phase of a multiply scattered field," USA (Radio propagation). 1962, 3, 231-240.
320. Beckman P., Sprizzidine A., *The scattering of electromagnetic waves from rough surface*, Pergamon, London, 1963.
321. Bellman R., *Adaptive Control Processes*, Princeton University Press, 1959.

322. Bello P. A., "Characterization of randomly time-variant linear channel," *IEEE Trans.* 1963, vol. VCS-11, 360-393.
323. Bendat J. S., *Principles and Applications of Random Noise Theory*, Y. Wiley, N.Y., 1958.
324. Bennett W. R., Curtis H. E., and Rice S. O., "Interchannel interference in FM and PM systems under noise-loading conditions," *BSTJ*, 1955, vol. 34, 3, 601-636.
325. Berlekamp E. R., *Algebraic coding*, McGraw-Hill, N.Y., 1968.
326. Blasbalg H., Freeman D., and Reeler R., "Random-Access Communications using Frequency Shifted PN (pseudo-noise) Signals," *IEEE Inter. Conv. Rec.*, 1964, part 6.
327. Bose R. C., Ray-Chaudhuri D. K., "On a class of error correcting codes," *Inform. and Control.*, 1960, vol. 3.
328. Braasch R. H., "The distribution of (n-m) terms for maximal length linear pseudo-random sequences," *IEEE Trans. Inf. Theory*, 1968, vol. IT-14, 14.
329. Brockbank R. A., Wass A. A., "Nonlinear distortion in transmission systems," Proc. IEE. 1945, vol. 92, part 3, 17, 45-56.
330. Broliu S. Y., Brown Y. M., "Companded delta modulation for telephony," *IEEE Trans. Commun.*, 1968, vol. COM-16, 1, 157-162.
331. Bullington K., Fraser I. M., "Engineering aspects of TASI," *BSTJ*, 1959, 3.
332. Burton F. R., "A survey of telephone speech-signal statistics and their significance in the choice of a p. cpm. companding law," *Proc. IEE*, 1962, vol. 109 B, 60-485.
333. Booker N. G. and Gordon E., "A theory of radio scattering in the troposphere," *Proc. IRE.*, 1950, Apr., vol., 38.
334. Cahn C. R., "Crosstalk due to finite limiting of frequency multiplexed signals," *Proc. IRE.*, 1960, Jan., vol. 48, 53-59.
335. Campbell M. R., Hoff L. E., and Ziemer R. E., "A large timebandwidth product signaling technique for nonwhite noise channels," *IEEE Trans. on Commun.*, 1976, vol. COM-24, 10.
336. Chau H. C., Taylor D. P., Haykin S. S., "Intelsat IECE3ITE," Third Internat. Conf. on Digital Satellite Communic., Japan, Kioto, November, 1975.
337. Chase D., "A combined coding and modulation approach for communication over dispersive channels," *IEEE Trans. Commun.*, 1973, vol. COM-24, 3
338. Chase D., "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inf. Theory*, 1972, Jan., vol. IT-18.
339. Chevillat P., "Fast sequential decoding and a new complete decoding algorithm," Technical report EE 7606-32-01, Illinois Institute of Technology, Chicago, 1976.
340. Cohen A. R., "Practical aspects of convolutional encoding with Viterbi decoding," *IEEE*, 1975, Internat. Conf. Comm., San-Francisco.
341. Chien R. T., Clayton V. E., Boudrenn P. E., and Locke R. R., "Error correction in a radio based data communication system," *IEEE Trans. Commun.*, 1975, 23, 4, 458-462.
342. Church R., "Tables of irreducible polynomials for the first power prime moduli," *Annals of mathematics*, 1935, vol. 36, 198-209.
343. Cook C. E., "Pulse-compression—key to more efficient radar transmission," *Proc. IRE.* 1960, Jan., vol. 48.
344. Cook C. E., Bernfeld M., *Radar signals*, Acad. Press, N.Y., 1967.
345. Costas I. P., "Poisson, Shannon and radio amateurs," *Proc. IRE*, 1959, Dec., vol. 47.
346. Counor D. Y., Braiuard R. C., "Intraframe coding for picture transmission," *IEEE*, 1972, 60, 8, 774-791.
347. Cramer H., Leadbetter M. R., *Stationary and Related Stochastic Processes*, J. Wiley, N.Y., 1967.

348. Crandall I. B., MacKenzie D., "Analysis of the energy distribution in speech," *Phys. Rev.* 1922, vol. 19, 221-231.
349. Cross T. G., "Intermodulation noise in FM systems due to transmission deviations and AM/PM conversion," *BSTJ*, 1966, 54, 12.
350. Caki F., *State-space Method for Control System*, Budapest, Akad. Kiado, 1976.
351. Cumminskey P., Jayant N. S. and Flanagan G. L., "Adaptive quantization in differential PCM coding of speech," *BSTJ*, 1973, 52, 7, 1105-1118.
352. Davies W. D. T., *Systemerkennung für adaptive Regelungen*, R. Oldenburg. 1973.
353. Davenport W. B., "An experimental study of speech-wave probability distributions," *JASA*, 1952, 24, 4, 390-399.
354. Davenport W. B., Root W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, N.Y., 1958.
355. Dawson C. H., Sklar H., "False addresses in random access system employing discrete time-frequency addressing," *IEEE Intern. Conv. Rec.*, 1964, 6.
356. Donaldson R. W., Douville R. I., "Analysis, subjective evaluation, optimization and comparison of the performance capabilities of PCM, DPCM, DM, AM and voice communication systems," *IEEE Trans. on Commun.*, 1964, COM-17, 4, 9, 421-431.
357. Dunn N. K., White S. P., "Statistical measurements on conversational speech", *JASA*, 1940, vol. 11, 1.
358. Dupont J. J., "Le systeme de radiotelegraphie multiplex T.O.R. (Telegraphy over Radio Circuits," *L'Onde Electrique*, 1954, 326.
359. Elias P., "Error-free coding," *IRE Trans.*, 1954, PGIT-4.
360. Elias P., *Coding for Two Noise Channels*, MIT, Mass, 1955.
361. Fano R., "A heuristic discussion of probabilistic decoding," *IEEE* Trans. 1963, IT-9, 2.
362. Fano R. M., *Transmission of Information*, MIT Press, N.Y., 1961.
363. Finger A., "Erzeugung pseudostochastischer Signale mit vorgebbaren Symbolhäufigkeiten," Nachrichtentechnik-Elektronik, 1973, 23, 4.
364. Finger A., "Ein Beitrag zur Erzeugung und Analyse diskreter pseudostochastischer Signale," Dissertation TU Dresden, 1973.
365. Forney G. D., Sr., *Concatenated Codes*, MIT Press., N.Y., 1966.
366. Foschini G. J., "Optimum direct detection for digital fiberoptic communication systems," *BSTJ*, 1975, 54, 8, 1389-1430.
367. Frank R. L., "Polyphase codes with good nonperiodic correlation properties," *IEEE Trans.*, 1963, IT-9, 8, 389-1430.
368. Fratta L., Gerla M., and Kleinrock L., "The flow deviation method. An approach to store-and-forward communication network design," Networks, 1973, 3, 98-133.
369. Fults G. L., Kleinrock L., "Adaptive routing techniques for store-forward computer-communication network," Proc. 1971 Int. Conf., Montreal, 1971.
370. Gabor D., *Theory of Communications*, IEEE, London, 93, 1946.
371. Gallager R., *Information Theory and Reliable Communication*, J. Wiley, N.Y., 1968.
372. Gallager R., "A minimum delay routing algorithm using distributed computation," *IEEE Trans. on Commun.*, 1977, COM-25, 73-85.
373. McGarby T. P., *Stochastic Systems and State Estimation*, J. Wiley, N.Y., 1974.
374. Gilbert E. N., "A comparison of signalling alphabets," *Bell System Techn.* J., 1952, Jan., vol. 31.
375. Golomb S. W., Scholtz R. A., "Generalized Barker sequences," *IEEE Trans.*, 1965, IT-11.
376. Golomb S. W., *Shift Register Sequences*, Holden-Day, San-Francisco, 1967.
377. Gordos G., Sallay Gy., "Computer assisted analysis of summed processes," Számitástudomanyi Konferencia, Székesfehérvár, 1973, maj.

378. Gordos G., "Statistical analysis of FDM systems with heterogeneous load," Tokai University — Budapest Technical University Joint Symposium, Tokai, 22-23, July, 1977.
379. Hagelbarger D. W., "Recurrent codes," *BSTJ*, 1959, 4.
380. Hajck J., Sidak Z., *Theory of Rank Test*, Acad. Sciences, Prague, 1967.
381. Hamer R., "Radio-frequency interference in multichannel telephony FM radio systems," *Proc. IEE*, 1961, part C, 108, 1, 75-89.
382. Hartley R. V. L., "Transmission of Information," Bell System Techn. J., 1928, July.
383. Hawkcs T. A., Simonpierri P. A., "Signal coding using a synchronous delta modulation," *IEEE Trans. on Commun.*, 1974, COM-22, 5, 729-730.
384. Hayes J. F., Shernan D. N., "A study of data multiplexing techniques and delay performance," *Bell. Syst. Techn. J.*, 1972, 51, 9.
385. Held M., Karp R. M., "A dynamic programming approach to sequencing problems," *L. Soc. Indust. and Appl. Math.*, 1962, 10, 1, 196-210.
386. Helstrom C. W., *Statistical Theory of Signal Detection*, Pergamon Press, London, 1960.
387. Hoequenghem A., "Codes correcteurs d'erreurs", *Chiffres*, vol. 2, 1959.
388. Hogt R. S., "Probability functions for the modulus and angle of the normal complex variety," *BSTJ*, 1947, 26, 2, 318.
389. Hollbrock B. D., Dixon J. T., "Load rating theory for multichannels amplifiers," *BSTJ*, 1939, 18, 1.
390. Huffman D. A., "The synthesis of linear sequential coding networks," in *Information Theory*, Acad. Press., N.Y., 1956.
391. Hullett J. L., "A modified receiver for digital optical fider transmission systems," *IEEE Trans. on Commun.*, 1975, COM-23, 12, 1518-1521.
392. Jayant N. S., "Digital coding of speech wave forms: RCM, DPCM, DH Quantizers," *Proc. IEEE*, 1974, 62, 5, 611-632.
393. Jayant N. S., "Adaptive quantization with one word memories," BSTJ, 1973, 52, 7, 1119-1145.
394. *Intelsat. Conf. on Satellite Communic. System. Techn.*, London, 1975, April.
395. *Intelsat. IEE Conf. on Digital Satellite Communication*, Editions Chiron, Paris, 1972.
396. Joshitaka Takasaki, "Optimal pulse formats for fiber optic digital communications," *IEEE Trans. on Commun.*, 1976, COM-24, 4, 404-412.
397. Instenst. J., "A class of constructive asymptotically good algebraic codes," *IEEE Trans.*, 1972, Sept., IT-18.
398. Kailath T., "Correlation detection of signals perturbed by random channel", Transact. *IRE Part Inf. Th.*, June, 1960.
399. Kailath T., "A view of three decades of linear filtering theory," *IEEE Trans.*, 1974, IT-20, 2.
400. Kalman R. E., "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, 1960, vol. 82D, 34-45.
401. Kalman R. E., Bucy R., "New results in linear filtering and prediction theory," *Trans. ASME J. Basic Eng.*, 1961, March, vol. 88D, 95-108.
402. Kalman R. E., Fald P. L., and Arbib M. A., *Topics in Mathematical System Theory*, McGraw-Hill, N.Y., 1969.
403. Kastenholz C. E., Birkemaier W. P., "A simultaneous information and channel sounding modulation technique for wideband channel," *IEEE Trans. Comm. Techn.*, 1965, 13, 2.
404. Kennedy R. S., *Fading Dispersive Communication Channels*, J. Wiley, N.Y., 1969.
405. Kerr I. H., Comberg G. R. A., Price W. L. and Solomonides C. M., "A simulation study of routing and flow control problems in a hierarchically connected packet switching network," Proc. 1976 ICCC, Toronto, 1976.
406. Khintchine A., "Korrelationstheorie der stationaren stochastischen Prozessen," *Math. Annalen*, 1934, 109.

407. Kikkert C. J., "Digital companding techniques," *IEEE Trans. on Commun.*, 1974, COM-22, 1, 75-78.
408. Kleinrock L., Lam S., "Packet switching in a multiaccess satellite channels," AFIPS Conf. Proc. 1973, vol. 42.
409. Kleinrock L., Tobagi F., "Packet switching in radio channels. Part 1," *IEEE Trans. on Commun.*, 1973, 23, 12.
410. Kleinrock L., Lam S., "Packet switching in a multiaccess broadcast channel performance," *IEEE Trans. on Commun.* 1976, 24, 8.
411. Kleinrock L., Tobagi F. A., "Carrier sense multiple access for packet switched radio channels," *PR Temporary Note*, Stanford Research Institute, Menio Park, Cal., vol. 91, Apr. 1974.
412. Kleinrock L., Kamoun F., "Data communications through large packet switching network," Proc. ICCC, Toronto, 1976.
413. Kleinrock L., *Queueing Systems. v. 1. Theory*, J. Wiley, N.Y., 1975.
414. Kleinrock L., *Queueing Systems. v. 2. Computer applications*, J. Wiley, N.Y., 1976.
415. Klauder I. R., Price A. C., Darlington S., and Albersheim W., "The theory and design of ship radars," *BSTJ*, 1960, 39, 1.
416. Kolmogoroff A., "Uber die analytischen Methoden in der Warscheinlichkeitsrechnung," *Math. Ann.*, 1931, B. 104, S. 415-458.
417. Kolmogorov A. N., "Sur l'interpolation et extrapolation de suits stationaires," *C.R. Acad. Sci.*, 1939, 208.
418. Kolmogorov A. N., "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans.*, 1956, Dec., vól. IT-2.
419. Konchecm A., Montev B., "Waiting lines in a system with polling," *J. As. Cimp. Mach.*, 1974, vol. 21.
420. Küpfüller K. *Die Systemtheorie der elektrischen Nachrichtenübertragung*, Hirzel, Stuttgart, 1952.
421. Lange F. H., *Korrelationselectronik*, Verlag Technik. Berlin, 1962.
422. Lange F. H., *Signale und Systeme. B. 3*, Technik, Berlin, 1971.
423. Lam S., "Packet switching in a multiaccess broadcast channel," *UCLA-ENG-7429*, March, 1974.
424. Lam S., Kleinrock L., "Packet switching in a multiaccess broadcast channel: dynamic control procedures," *IEEE Trans. on Commun.*, 1975, COM-23, 891-904.
425. Baghdady E. J. (Ed.), *Lectures on Communication System Theory*, 3d ed, McGraw-Hill, N.Y., 1961.
426. Leman E., *Testing Statistical Hypothess*, J. Wiley, N.Y., 1959.
427. Levin B. R., "Asymptotically optimal techniques for testing the nonstationarity of stochastic process," European meeting of statisticians, Budapest, 1972.
428. Levin B. R., "Asymptotische Methoden in der statistischen Nachrichtentechnik," *Nachrichtentechnik und Electronik*, 1974, 3.
429. Loondes C. T., (Ed.), *Theory and Application of Kalman Filtering*. NATO Advanced Group for Aerospace, AGARDO graph. 139, 1970, Feb.
430. Levy P., *Processus stochastiques et mouvement Brownien*, Paris, 1965.
431. Lindholm I. H., "An analysis of the pseudo-randomness properties of subsequences of long m-sequences," *IEEE Trans. Inf. Theory*, 1968, IT-14, 4.
432. Lumb D. R., "Test and preliminary flight results of the sequential decoding of convolutionally encoded data from Pioneer-9", IEEE Int. Communications Conf. Rec., Bouldek, Colo., 1969.
433. Lyghounis E., Poretti J., and Monti G., "Speech interpolation in digita transmission system", *IEEE Trans. on Commun.*, 1974, 9.
434. Maritsas D. G., Hartley M. G., "A case of study of a versatile generator of repeatable non-Poisson sequences of pseudorandom pulses," *IEEE Trans.* 1970, vol. C-19, 9, 812-817; 10, 924-934.

435. Massy J. L., *Threshold Decoding*, MIT Press, N.Y., 1963.
436. Hax J., "Quantizing for minimum distortion," *IRE Trans.*, 1960, IT-16, 3, 7-12.
437. Mazo I. E., "On optical data communication via direct detection of light pulses," *BSTJ*, 1976, vol. 55, 3.
438. Medgyessy P., *Decomposition of superpositions of density functions and discrete distributions*, Akad. Klado, Budapest, 1977.
439. Medhurest R. G., Hicks E. N., and Grosset W., "Distortion in frequency division multiplex FM systems due to an interfering carrier," *Proc. IEE*, 1958, part C, 105, 5, 282-292.
440. Metalfe R. M., "Steady-state analysis of a slotted and controlled ALOHA system with blocking," Proc. of the Sixth Hawaii Int. Conf. on System Sciences, 1973, Western Periodicals Comp.
441. Middleton D., An Introduction to Statistical Communication Theory, McGraw-Hill, N.Y., 1960.
442. Middleton D. A., "Statistical theory of reverberation and similar first-order scattered field," *IEE Trans.*, 1963, July, IT-13, 372-414.
443. Mitru D., "Mathematical analysis of an adaptive quantizer," *BSTJ*, 1974, 53, 5, 867-898.
444. Nakagami M., *The m-Distribution of a General Formula of Intensity Wave Propagation*, N.Y., 1960.
445. Nakagami M., On the intensity distribution

$$\frac{2R}{\sqrt{\alpha\beta}} \exp\left[-\frac{R^2}{2}\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)\right]\times I_0\left[\frac{R^2}{2}\left(\frac{1}{\beta}-\frac{1}{\alpha}\right)\right]$$

and its application to signal statistics," Radio science J. of research NES (USNC-URSI), 1964, September, vol. 68D.
446. Neuvo Y., Walter H. K., "Analysis and digital realization of a pseudorandom gaussian and impulsive noise source," *IEEE Trans.*, 1975, COM-23, 9, 849-858.
447. Narth D. O., "Analysis of factors which determine signal-to-noise discrimination in pulsed carrier systems," Rep. PTR-6C, RCA, Princeton, 1943, reprinted in Proc. IRE, 1963, vol. 51, July.
448. Nyquist N., "Certain factors effecting telegraph speed," *BSTJ*, 1924, 4, 324 (see also: "Certain topics in telegraph transmission theory"), *AIEE Trans.*, 1928, Apr., vol. 47, 617.
449. Pangratz H., "Schneller dezimaler Pseudozufallsgenerator mit guten Korrelationseigenschaften," *Frequenz*, 1976, 30, 8, 204-209.
450. Patrick E. A., *Fundamentals of Pattern Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1972.
451. Personic S. D., "Receiver design for digital fiber optic communication systems," *BSTJ*, 1973, 52, 6, 843-886.
452. Peschel M., *Anwendung statistischer Methoden in der Regelungstechnik. Statistische Modellbildung. Reine Automatisierungstechnik*, Verlag Technik, Berlin, 1972.
453. Peterson W. W., *Error-correcting Codes*, MIT Press, N.Y., 1961.
454. Pierce I. A., "An introduction to Loran", Proc. IRE, 1946, 34, 5.
455. Price P., Green P. A., "Communication technique for multipath channels," *Proc. IRE*, 1958, 9, 555-573.
456. Rice S. O., "Mathematical analysis of random noise," *BSTJ*, 1944, vol. 23, 283-332; 1945, vol. 24.
457. Rice S. O., "Distortion produced in a noise modulated FM signal by nonlinear attenuation and phase shift," *BSTJ*, 1957, vol. 36, 879-889.
458. Rice S. O., "Second and third order modulation terms in the distortion produced when noise modulated FM waves are filtered," *BSTJ*, 1969, 57, 1, 879-889.

459. Richards D. L., "Statistical properties of speech signals," *Proc. IEE*, 1964 111, 5, 87-141.
460. Rihaczek A. W., "Radar resolution properties of pulse trains," *Proc. IEEE*, 1964, 52, 941-949.
461. Roberts L. G., "Multiple computer networks and intercomputer communication," Proc. ACM Symp. on Operating Systems, Gatlinbourg, Ten., 1967.
462. Roberts L. G., "Dynamic allocation of satellite capacity through packet reservation," *AFIPS Conf. Proc.*, vol. 42, National Computer Conference 1973.
463. Rudin H., "An introduction to flow control," *Proc. ICCC*, 1976, Toronto.
464. Rudin H., "On routing and "Delta Routing": A taxonomy and performance comparison of techniques for packet-switched networks," *IEEE Trans. on Commun.*, 1976, vol. COM-24.
465. Ruthroff C. L., "A mechanism for direct adjacent channel interference," *Proc. IRE* 1961, 49, 49.
466. Sage A. P., Melsa J. S., *Estimation Theory with Application to Communication and Control.*, McGraw-Hill, N.Y., 1972.
467. Sage A. P., White C. W., *Optimum System Control*, Second ed., Prentice-Hall, N.Y., 1977.
468. Samoylenko S. I., "Man-computer problem solving in computer communications networks," Proc. of the Second International Conference on Computer Communications. Stockholm, 1974.
469. Schindler H. K., "Delta-modulation," *IEEE Spectrum*, 1970, Oct.
470. Schwarz W., "Verhaltensmodelle zeitdiskreter stochastischer systeme," Dissertation B, Techniche Universität Dresden, 1976.
471. Schwartz M., Cheung C. K., "The gradient projection algorithm for multiple routing in message-switched networks," *IEEE Trans. on Commun.*, 1976, COM-24, 449-456.
472. Segall A., "The modeling of adaptive routing in data-communication networks," *IEEE Trans. on Commun.*, 1977, COM-25.
473. Seidler J., *Systemy przesylania informacji cyfrowych*, WNT, Warszawa, 1976.
474. Seidler J., "Analiza i synteza sieci lacznosci dla systemów teleinformatycznych, WNT, Warszawa 1978.
475. Seidler J., Konorski J., "Average length of packet route in a network with mixed deterministic random routing," *Bull. Ac. Sci. Pol.*, Sec. IV, 1978.
476. Shakin V. V., Breuer P., "Adaptive least-squares spline fitting the vectorial signals," *Proc. Conf. Digital Processing*, Florence, 1975.
477. Shamblin J. F., Stevens G. T., *Operation Research*, McGraw-Hill, Inc., N.Y., 1974.
478. Shannon C., "A mathematical theory of communication," BSTJ, 1948, 27, 3, 4.
479. Shannon C., "Certain results in coding theory for noisy channels," *Inform. and Control*, 1, Sept., 1957.
480. Shannon C., "Probability of error optimal codes in a gaussian channel," *BSTJ*, 1959, vol. 38.
481. Shannon C., Gallager R. G., Berlekamp E. R., "Lower bound to error probability for coding on discrete memoryless channels," *Inform. and Control*, 1967, vol. 10.
482. Shneider K. S., Orr R. S., "Aperiodic correlation constraints on large binary sequence sets," *IEEE Trans. Inf. Theory*, 1975, IT-21, 1.
483. Silk D. J., "Routing doctrines and their implementation in message-switching network," *Proc. Inst. Elec. Eng.*, 1969, vol. 116.
484. Snyder D. L., "The state-variable approach to continuous estimation with application to analog communication theory," *MIT Res. Monogr.*, 1969.
485. Sobolewski J. S., Pyne W. H., "Pseudonoise with arbitrary amplitude distribution," *IEEE Trans.*, 1972, C-21, 4, 337-352.

486. Stiffler S. J., *Theory of Synchronous Communication*, Prentice-Hall, New Jersey, 1971.
487. Steele R., *Delta-modulation Systems*, Pentech., 1975.
488. Sunde E. D., "Intermodulation distortion in multicarrier FM systems," *IEEE International Convention Record*, 1965, part 2, 130-146.
489. Takasaki J., Tanaka M., "Optical pulse formats for fiber optic digital communications, *IEEE Trans. on Commun.*, 1976, COM-24, 4, 404-412.
490. Taki Y., Miyakawa H., Akiyama M., "Recent progress in communication theory in Japan," *IEEE Trans. on Commun.*, 1972, COM-20, 4, 696-707.
491. Turin G. L., "An introduction to matched filters," *IRE Trans.*, 1960, vol. IT-6.
492. Turin G. L., "Communication through noisy, random-multipath channels," *IRE National Convention Record*, 1956, ptr. 3.
493. Turin G. L., "Error probabilities for binary symmetrical ideal reception through nonselective slow fading noise," *Proc. IRE*, 1958, Sept. 46, 9, 1603-1619.
494. Tzafestas S. G., Nightingale I. H., "Optimum filtering, smoothing and prediction in linear distributed parametre systems," *Proc. IEE*, 1968, Aug., vol. 115, 1200-1212.
495. Van Trees H. L., *Detection, Estimation and Modulation Theory*, J. Wiley, N.Y., 1969.
496. Viterbi A. J., *Principles of Coherent Communication*, McGraw-Hill, N.Y., 1966.
497. Viterbi A. J., "A new coding technique for asynchronous multiple access communication," *IEEE Trans. Com. Tech.*, 1971, COM-19, 5.
498. Webb P. R. W., "Military satellite communications using small earth terminals," *IEEE Aerosp. Electr. Syst.*, 1974, AES-10, 3.
499. Wiener N., "Generalized harmonic analysis," *Acta Mathematica*, 1930, vol. 5.
500. Wiener, N., *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, J. Wiley, N.Y., 1949.
501. Wiener N., *I am a Mathematician*, Doubleday Inc., N.Y., 1956.
502. Westcott R. J., "Investigation of multiple FM/FDM carriers through a satellite TWT operating near to saturation," *Proc. IEEE*, 1967, 114, 6, 726-739.
503. Wolfowits J., *Coding Theorems of Information Theory*, Springer, Berlin, 1964.
504. Woodward P. M., *Probability and Information Theory with Application to Radar*, Pergamon Press, London, 1953.
505. Wozencraft J. M., "Sequential decoding for reliable communication," *IRE Convention Record*, 1957, 2.
506. Wozencraft J. M., Jacobs I. M., *Principles of Communication Engineering*, J. Wiley, N.Y., 1965.
507. Wunsch G., *Systemtheorie*, Geest & Portig, Leipzig, 1975.
508. Zadeh L. A., Desoer C. A., *Linear System Theory*, McGraw-Hill, N.Y., 1967.
509. Zettebery L. H., Uddenfeldt I., "Adaptive delta modulation with delayed decision," *IEEE Trans. on Commun.*, 1974, COM-22, 9, 1195-1198.
510. Zierler N., "Linear recurring sequences," *J. Soc. Indust. Appl. Math.*, 1959, 7, 1, 34-48.

Author Index

Subject Index

D

E

F

G

H

K

L

M

N

To the reader

Mir Publishers welcome your comments on the content, translation, and design of the book.
We would also be pleased to receive any suggestions you care to make about our future publications.
Our address is:
USSR, 129820, Moscow, I-110, Pervy Rizhsky Pereulok, 2, Mir Publishers.

Printed in the Union of Soviet Socialist Republics